Dual-luciferase assay comparison of twenty mouse genes' 3′UTRs to a conditional knockup flex-cassette

Mark Cary Cowlishaw

Institute of Biotechnology
Faculty of Biological and Environmental Sciences
Degree Programme in Biology
Genetics Major

University of Helsinki
Finland

Master's Degree Thesis

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta – Fakultet – Faculty	Laitos – Institution – Department
The Faculty of Biological and Environmental Sciences	Degree Programme in Biology

Tekijä – Författare – Author

Mark Cary Cowlishaw

Työn nimi – Arbetets titel – Title

Dual-luciferase assay comparison of twenty mouse genes' 3′UTRs to a conditional knockup flex-cassette

Oppiaine – Läroämne – Subject

Genetics

Työn laji – Arbetets art – Level	Aika – Datum – Month and year	Sivumäärä – Sidoantal – Number of pages
Master's Degree	September 2020	75

Tiivistelmä – Referat – Abstract

Upregulation of specific helpful proteins represents a possible method for preventing or treating human diseases. Endogenous upregulation (knockup) is the increase of a gene's expression only in cells in which it is already expressed, thus avoiding physiologically abnormal spatiotemporal patterning.

A gene's three prime untranslated region (3′UTR) affects protein expression through stability regulation of RNA already transcribed, which suggests 3′UTR modification as a viable route for endogenous upregulation. Mammalian model organisms can be generated in order to test the effects of different 3′UTR modifications, but at great cost of time, effort, and money. If able to predict in advance with an in vitro assay whether an in vivo modification would cause a desirable or undesirable change, these costs could be substantially reduced.

In this thesis project, an in vitro assay was used to compare the protein expression influence of twenty neurodegeneration-relevant mouse genes' 3′UTRs to that of a flip-excision cassette (flex-cassette) previously used for in vivo conditional knockup. The assay used was the Promega Dual-Luciferase Reporter Assay, in which plasmids expressing Renilla and Firefly luciferase as reporter and internal control are co-transfected into in vitro cells, then each luciferase's expression measured with its respective substrate and a luminometer. Transfections were carried out in three-well replicates and on multiple days.

The aims of the project were the evaluation of the assay's ability to predict in vivo results, the suggestion of 3′UTRs which could be upregulated in vivo by the conditional knockup flex-cassette, and the identification of any trends in 3′UTR-based protein expression influence according to gene function.

A number of gene 3′UTRs were identified which were either candidates for flex-cassette upregulation or candidates for use in the flex-cassette to upregulate other genes. However, the flex-cassette's in vitro results were only partially consistent with its previous in vivo results. Specifically, the lox sites in the flex-cassette was observed to lower expression level to a degree not observed in vivo. Additionally, in the course of the project a number of possible workflow improvements were identified, for which suggestions have been made in the text. As such, this in vitro approach requires further study in order to determine suitability for prediction of in vivo 3′UTR behaviour.

Avainsanat – Nyckelord – Keywords

protein expression, conditional knockup, flex-cassette, in vitro, luciferase assay

Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors

Associate Professor Jaan-Olle Andressoo, PhD

Säilytyspaikka – Förvaringställe – Where deposited

University of Helsinki Digital Theses E-Thesis service

Muita tietoja – Övriga uppgifter – Additional information

Table of Contents

1. Abbreviations used....1
2. Acknowledgements....2
3. Introduction....3
    3.1. Treatment design for cell dysfunction diseases....3
        3.1.1. Protein- versus DNA-level treatment....3
        3.1.2. Protein sequence alteration versus protein expression alteration....3
        3.1.3. Desirable protein expression changes....4
        3.1.4. Types of expression alteration possible....4
        3.1.5. Three prime untranslated region (3′UTR) knockup....5
    3.2. Non-human testing....7
        3.2.1. In vivo modelling with a conditional knockup flex-cassette....7
        3.2.2. In vitro modelling with a dual-luciferase assay....8
4. Aims....12
5. Materials and Methods....13
    5.1. Plasmid amplification....13
    5.2. Reporter plasmid construction....13
        5.2.1. Primer design....13
        5.2.2. PCR of 3′UTR sequence inserts....14
        5.2.3. Restriction digestion for cloning....15
        5.2.4. Electrophoresis and gel purification....16
        5.2.5. Ligation of insert and backbone....16
    5.3. Co-transfection....17
    5.4. Measurement recording....17
    5.5. Statistical analysis....18
        5.5.1. Testing of samples for homogeneity of variance....18
        5.5.2. Testing of samples for normality....18
    5.6. Day-adjustment normalisation of recorded data points....18
    5.7. Graphical representation of data points....19
6. Results and Discussion....20
    6.1. Differing plasmid DNA degradation and the implementation of ‘batches’....20
    6.2. Statistical analysis....22
        6.2.1. Initial intentions....22
        6.2.2. Insufficient evidence for homogeneity of samples' population variance....23
        6.2.3. Insufficient evidence for samples being normally-distributed....26
    6.3. All measurements in unadjusted and day-adjusted forms....27
    6.4. Flex-cassette conditional knockup testing....28
        6.4.1. Initial test....28
        6.4.2. Readthrough-influence test....31
        6.4.3. Lox-influence test....33
    6.5. Comparison of different co-transfection ratios....34
    6.6. Candidates for bGHpA flex-cassette conditional knockup....36
    6.7. Alternatives to bGHpA for flex-cassette 3′UTR replacement....38
    6.8. Possible workflow improvements for future experiments....40
        6.8.1. Mouse cell line....40
        6.8.2. Plasmid condition and concentration measurement....40
        6.8.3. RNA transcript sequence verification....41
        6.8.4. Flex-cassette internal sequence....41
        6.8.5. Statistical analysis method....42
7. Conclusions....43
8. References....44
9. Appendices....49
    9.1. Appendix A: Measurement day and 3′UTR type abbreviations and identities....49
    9.2. Appendix B: Non-genomic template sequences....52
        9.2.1. Flex-cassette template....52
        9.2.2. WPRE+hGH template....53
    9.3. Appendix C: Primers used....53
    9.4. Appendix D: Raw and processed measurement data....55

1. Abbreviations used

Throughout this text abbreviations are used for the measurement days and the 3′UTR types used; information on these abbreviations can be found in Appendix A.

In gene abbreviations such as bGH (for bovine growth hormone), an initial lowercase ‘m’, ‘b’, or ‘h’ indicates a mouse, bovine, or human origin respectively.

°C	degrees Celsius
3′	three prime
3′UTR	three prime untranslated region
5′	five prime
ANOVA	analysis of variance
BAC	bacterial artificial chromosome
bGHpA	bovine growth hormone polyadenylation signal sequence
bp	base pair/s (typically of DNA)
CDF	cumulative distribution function
control:reporter	co-transfection molar ratio of Firefly-lucferase-expressing internal control plasmid to Renilla-luciferase-expressing reporter plasmid
CPSF	cleavage and polyadenylation specificity factor (protein complex)
CstF	cleavage stimulation factor (protein complex)
DNA	deoxyribonucleic acid
flex-cassette	flip-excision cassette (Schnütgen et al., 2003 and 2005), for which Cre recombinase irreversibly flips a sequence's orientation
kb or kbp	kilobase/s or kilobase pair/s (one kilobase equivalent to 1000 bases)
ml	millilitre/s
mRNA	messenger RNA
N/A	not applicable
NCBI	National Center for Biotechnology Information
n.d.	no date
ng	nanogram/s
n.n.	no name
RNA	ribonucleic acid
SOC medium	super optimal broth with catabolite repression
µg	microgram/s
µg/ml	micrograms per millilitre
µl	microlitre/s
UV	ultraviolet
PCR	polymerase chain reaction
TAE	Tris-acetate-EDTA (buffer)
Tm	theoretical temperature at which half of a primer's molecules are annealed and half unannealed, used as the annealing temperature in PCR
V	volts

2. Acknowledgements

This Master's thesis was carried out in the University of Helsinki while being a member of the Andressoo research group. I owe thanks to all Andressoo group members for exchange of ideas and help learning unfamiliar laboratory protocols. I further owe thanks to Anmol Kumar, a former member who who left before my time in the Andressoo research group, who previously constructed one of the plasmids used in this project

Special acknowledgements for those without whom this thesis would not have been possible:

Thesis supervisor
Associate Professor Jaan-Olle Andressoo, PhD

Thesis project supervision
Kärt Mätlik, PhD

Student advisor
University Lecturer Pekka Heino, PhD

Thesis writing feedback
University Lecturer Janne Ravantti, PhD

Life support
Emi Cowlishaw, my wife

3. Introduction

3.1. Treatment design for cell dysfunction diseases

3.1.1. Protein- versus DNA-level treatment

A cell's behaviour is controlled by its genetic and epigenetic information. Human protein-coding genes produce tens of thousands of proteins which function in regulated coordination to keep their cell in working order (Hahn and Wray, 2002). Disruptions to either protein coding sequence or protein expression regulation can give rise to a multitude of different pathologies, and pharmaceutical compounds or other tools which directly interact with proteins—protein-level methods—must be designed according to the structure and behaviour of each protein target (Chen et al., 2012).

By contrast, tools for genetic or epigenetic manipulation which can be used for any and all genes—DNA-level methods—do not have this limitation. Instead of expending money and time to develop and test an effective protein-level treatment for only one disease, then begin again from scratch for the next disease, a tool which affects protein generation requires development in theory only once before it can be used in testing for all genes, as all genes in a genome are encoded with the same storage format (Gurevich and Gurevich, 2015).

Further, if treatment is protein-level in nature and does not originate from the DNA inside target cells, specificity is difficult. A protein-level treatment will affect all susceptible cells exposed to it, as in when dopaminergic axons have grown towards a glial cell line-derived neurotrophic factor (GDNF) injection site rather than towards their physiological targets (Ibáñez and Andressoo, 2017). By contrast, working at the DNA level can allow conditional outcomes through expression regulation, without adversely affecting unrelated cells.

3.1.2. Protein sequence alteration versus protein expression alteration

Ultimately, manipulation of both protein sequence and protein expression are necessary for maintained cell health, both develop problems in the face of adverse conditions or the passage of time (Milholland et al., 2017). Dysfunctionally-expressed functional proteins and functionally-expressed dysfunctional proteins are both inadequate for cell functionality.

Both manipulation approaches lend themselves soonest to the treatment of inherited genetic disorders, in which all of a person's cells possess the same pathogenic allele and there exist people possessing non-pathogenic versions of that allele which can be used as templates for all accessible cells simultaneously. New mutations and epigenetic changes such as from aging are instead different for each affected cell, requiring either logistically infeasible cell-by-cell treatment or else beneficial alterations to all cells for prevention or mitigation of categories of problems which have similar effects.

If attempting beneficial alterations through protein sequence alteration, lacking a human template for the alteration desired, the challenge of designing the new protein structure according to its desired behaviour is even harder than that for pharmacuetical design, due to the complexity involved in protein-folding and protein domain interaction.

By contrast, expression alterations make use of existing protein structures and interactions while only modulating their effects through dosage, aiming for qualitative improvements with few uncertain factors—‘low-hanging fruit’ to be picked first.

3.1.3. Desirable protein expression changes

As part of the Andressoo research group's focus on neurodegeneration, attention has been paid to proteins in the categories of aggregation, anti-aggregation, and neuron survival. Aggregation proteins like amyloid beta for Alzheimer's disease are expressed in the first place because of their genes' beneficial or necessary functionality (Corrigan et al., 2012), but in a failure of autophagy can form harmful aggregates (McCray and Taylor, 2008). Anti-aggregation proteins are those which in healthy cells prevent the formation of pathogenic aggregates (Gao et al., 2015). Neuron survival proteins are those which stop neurons from dying despite not affecting their pathogenic aggregates; for GDNF, which promotes dopaminergic neuron survival, it is hoped that higher levels of GDNF could prolong or otherwise maintain neuron survival in Parkinson's disease patients (Kumar et al., 2015).

As with anti-aggregation proteins, increased thoroughness of DNA maintenance proteins offers to reduce the rate of new mutations in the first place. Particularly in the area of DNA double-strand break repair (Tian et al., 2017), there is correlation between repair gene expression and organism lifespan.

A common trend is of desirable beneficial alterations being in the form of increased expression of genes with beneficial functionality, with decreased expression only desired in cases where expression is already pathogenically abnormal.

3.1.4. Types of expression alteration possible

Within a given cell, for a protein's dosage rather than its localisation or other behaviour there are only four types of expression alteration possible: making non-zero expression zero, making zero expression non-zero, decreasing expression without eliminating it entirely, and increasing expression. These can be thought of as ‘1-to-0’, ‘0-to-1’, ‘less’, and ‘more’.

Non-zero expression being made zero (‘1-to-0’) is referred to as a knockout, and can be accomplished by excising from the genome either an entire gene or a critical component of that gene.

Making zero expression non-zero (‘0-to-1’) is not referred to symmetrically as a knockin, as the term knockin is already used for any targeted genomic insertion or replacement, unrelated to the resulting effect on protein expression. Instead, the term exogenous overexpression can be used for this, in which a protein is expressed from a plasmid or genomically-inserted coding sequence copy separately from its native allele, or else expressed from the native copy in cell types which would otherwise not express it. Exogenous overexpression can be used for either 0-to-1 expression or adding expression to expression already present; however, different overexpression contexts can result in very different expression patterns (Maskri et al., 2004). Exogenous overexpression is thus less reliable for reproducible treatment testing or gene-function interrogation than the other expression alteration types, which alter already-present expression from a native allele endogenously, without changing its spatiotemporal patterning.

Expression being decreased (‘less’) is referred to as a knockdown, and is accomplished for instance through the introduction of morpholinos which temporarily block messenger RNA (mRNA) translation.

Expression being increased (‘more’) by symmetry can be referred to as a knockup, and is an increasing of native expression only where already present. As such, this rules out any approach which alters a gene's promoter or other upstream regulatory sequence which controls in which cell types the protein is expressed.

3.1.5. Three prime untranslated region (3′UTR) knockup

Regulatory sequences at the five prime (5′) end of a gene, influenced by related enhancers, control whether it is transcribed or not in each cell. As transcription progresses from the mRNA's 5′ end to its three prime (3′) end, the primary transcript undergoes capping, splicing, and cleavage-and-polyadenylation; this produces a mature mRNA with a cap, a 5′ untranslated region, a protein-coding translated region beginning with a start codon and ending with a stop codon, a three prime untranslated region (3′UTR), and a poly(A) tail (Fig. 1).

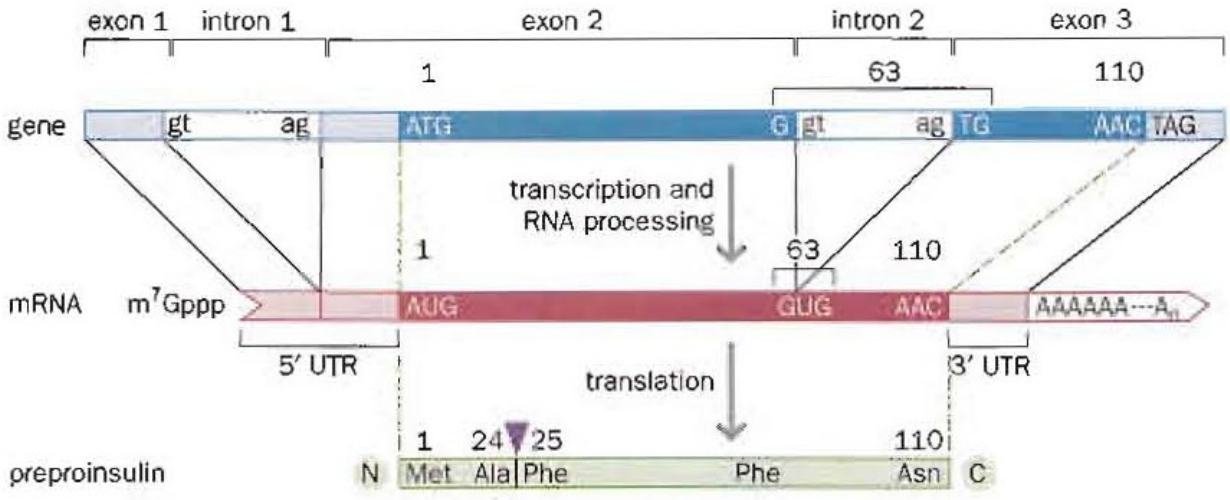

Figure 1. An example of messenger RNA (mRNA) structure, adapted and modified from "Insulin synthesis involves multiple post-translational cleavages of polypeptide precursors", Figure 1.26 of the fourth edition of *Human Molecular Genetics* (Strachan and Read, 2010). The mRNA three prime untranslated region (3′UTR) is transcribed from genomic DNA, but not translated into an amino acid sequence.

Known 3′UTRs are of 19 to 19,142 nucleotides long (Mignone, 2002). Transcription begins and proceeds from the gene's promoter at the 5′ end, while the 3′UTR takes a leading role posttranscriptionally by controlling poly(A) tail initial length and degradation rate through binding sites for microRNAs and RNA-binding proteins (Jalkanen et al., 2014). Binding sites for microRNAs for example are largely constrained to the 3′UTR (Shivdasani et al. 2006; Gu et al., 2009), and allow posttranscriptional regulation of different mRNAs according to which microRNAs each cell type expresses (Fig. 2). As an mRNA with a fully-degraded poly(A) tail is decapped and destroyed, this determines how long on average an mRNA molecule survives, and thus how many protein molecules on average are translated from each mRNA molecule.

This suggests the 3′UTR as a way to accomplish gene knockup, for example through targeted ablation of binding sites for microRNAs known to be highly expressed in the cell type or types of interest, ablation of other sequences known to decrease mRNA half-life, or instead by insertion of sequences known to increase mRNA half-life such as the cytoplasmic polyadenylation element.

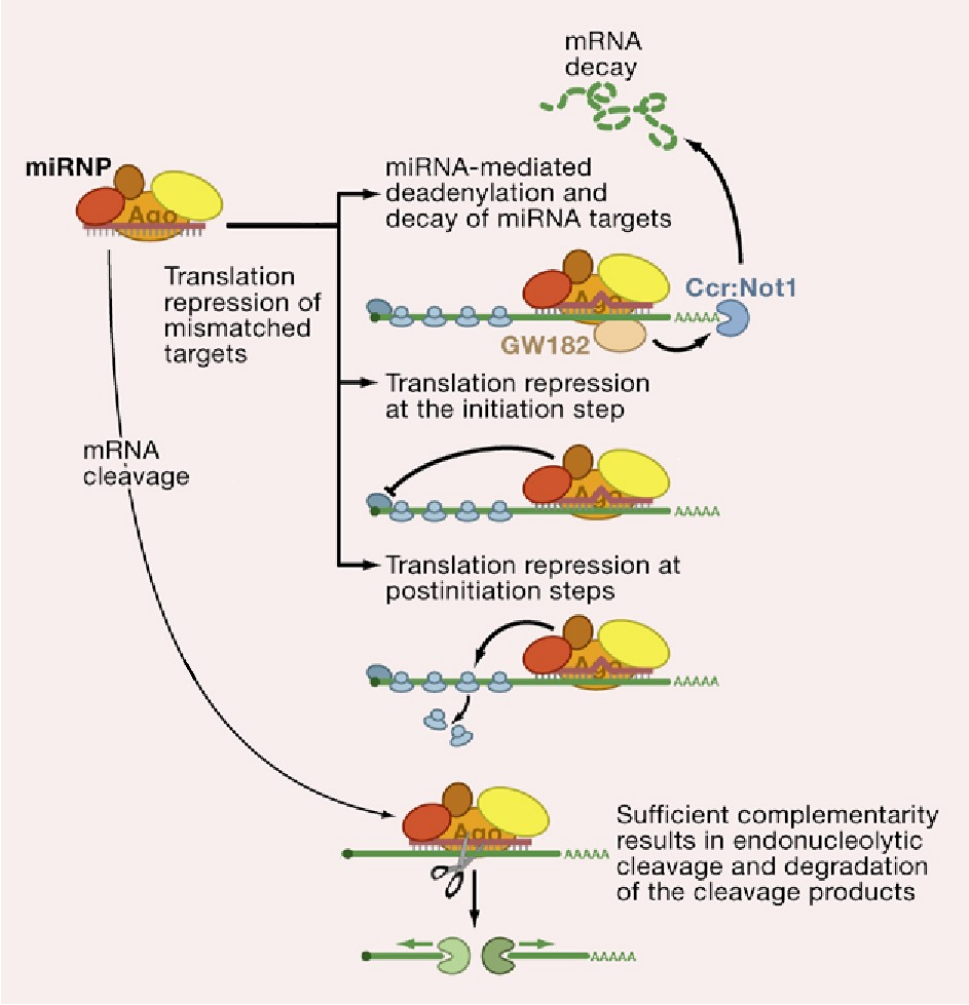

Figure 2. Expression downregulation through the 3′UTR, adapted and modified from "Posttranscriptional Gene Silencing" (Pressman et al., 2007). This figure depicts microRNAs (miRNAs) and their role in reduction of messenger RNA (mRNA) translation. miRNPs are microRNA ribonucleoprotein complexes.

3.2. Non-human testing

3.2.1. In vivo modelling with a conditional knockup flex-cassette

Before implementation as a human medical treatment, the process must first undergo non-human testing: even for humans facing the prospect of brain death, treatments attempted can thus be limited to those which have offered satisfactory results in non-human models, following the three Rs of **R**eplacement, **R**eduction, and **R**efinement propounded in 1959 by Russell and Burch (Emerson, 2010).

Rather than using a treatment method for adult animals, genetically-modified models were grown from embryonic stem cell clones. While treatment methods in adult humans might achieve varying levels of cell-affecting efficiency, a modification present in all cells allows theoretically 100% efficiency, which is desirable both for proof-of-concept assessing whether a treatment has hope of a positive result and for assessing whether that extreme 100% efficiency can bring about observable negative effects. It can also be noted that hypothetical future technological advances in adult-treatment efficiency could allow application of information about high-efficiency treatment results that might not have been obtainable if using a less-efficient technology to test the effects.

For initial human-relevant in vivo modelling, the mouse is currently the most favoured organism due to that it is mammalian, readily available, and has a number of well-established genetic modification methods (Nguyen and Xu, 2008; Ellenbroek and Youn, 2016). That said, while the time cost of growing few or many mouse models to adulthood is similar, the monetary cost of feeding and other care is not small, and varies proportionately with the number of mice being cared for. For a proof-of-concept in vivo test, a multitude of mice with many different alterations to each gene-of-interest's 3′UTR is monetarily and ethically wasteful: if critical information about the feasibility of 3′UTR alteration can be obtained through a less costly workflow, then again by the three Rs those excess resources should be used where more worthwhile.

When testing knockup of many genes, a single reliable alteration that could be applied in an assembly-line manner would also save time that would otherwise go into separately interrogating each gene's 3′UTR for individual modification.

For proof of concept experiments, the Andressoo research group tested the potential of 3’UTR alteration gene knockup by removing all native 3′UTR regulatory elements. This was done through the insertion of a different 3′UTR, thought to impart a long mRNA half-life, at the beginning of the native 3′UTR. After the inserted 3′UTR's cleavage-and-polyadenylation, all the native 3′UTR regulatory elements are thus in the cleaved-away downstream portion. The replacement 3′UTR used is the bovine growth hormone polyadenylation signal sequence (bGHpA), which displayed the highest effect on expression in a paper which tested the effects of different 3′UTRs on protein expression in mouse genetically-modified embryonic stem cells (Kakoki et al., 2004).

The result of this was the GDNF hypermorphic mouse, which exhibited increased GDNF mRNA levels and protein expression in the striatum, but also defective kidneys when homozygous due to the role of GDNF in kidney development (Kumar et al., 2015).

Building on this, the next form of gene knockup that the Andressoo research group has tested is a bGHpA flex-cassette, for which backward-orientation bGHpA is irreversibly excision-flipped to forward-orientation only in the presence of Cre recombinase (Schnütgen et al., 2003 and 2005). By crossing these conditional knockup mice with specific Cre recombinase expressing mouse lines, or by introducing Cre recombinase with viral delivery, upregulation can be induced at different developmental stages and/or different tissues and cell types (Mätlik et al., 2019).

In GDNF conditional knockup mice with ubiquitously-expressed Cre, the kidney size was reduced as in the GDNF hypermorphic mouse. However, in GDNF conditional knockup mice with Cre expressed from a Nestin promoter, GDNF knockup was restricted to the central nervous system. In both cases GDNF mRNA levels were elevated (Fig. 3), but much more so in the kidney, thought to be due to cell-type-dependent differences in GDNF wild-type mRNA half-life.

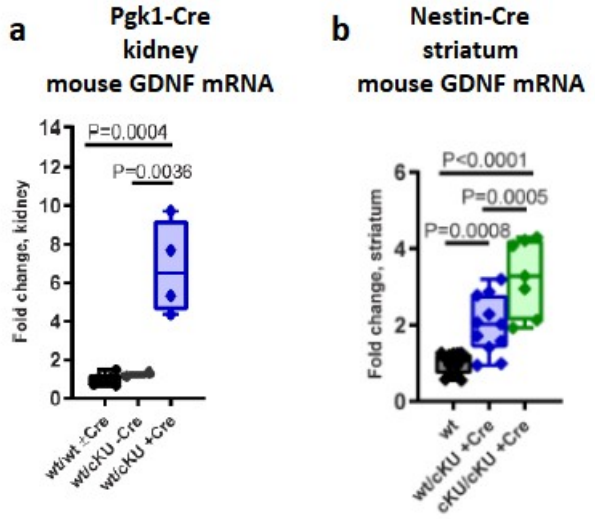

Figure 3. GDNF conditional knockup mRNA increases, adapted and modified from **(a) "**Endogenous GDNF in the kidney" and **(b) "**Endogenous GDNF in the brain of adult $Gdnf^{cKU}$;Nestin-Cre mice", Figure 2a and Figure 3a respectively of an Andressoo research group paper (Mätlik et al., 2019). These are the in vivo results of bGHpA flex-cassette conditional knockup in the mouse GDNF 3′UTR. wt is wild type, cKU is the conditional knockup allele, and +Cre indicates the presence of the Cre recombinase which excision-flips the flex-cassette's bGHpA sequence to its active forward-orientation. Pgk1-Cre is Cre recombinase expressed from a phosphoglycerate kinase 1 promoter, and Nestin-Cre is Cre recombinase expressed from a Nestin promoter.

3.2.2. In vitro modelling with a dual-luciferase assay

Though the aim has been to achieve a large effect by complete replacement with a long-half-life 3′UTR rather than incrementally increase the existing 3′UTR's half-life, the effect could be knockdown rather than knockup if replacing a 3′UTR which granted a longer half-life in a given cell type than the replacement 3′UTR.

As the protein expression of genes in vivo is affected by 5′ transcriptional regulation and not only by 3′UTR posttranscriptional regulation, the effect of the flex-cassette cannot be predicted from wild-type in vivo expression levels. Again following the three Rs, it would be helpful to be able to test the effect of a flex-cassette on a 3′UTR's half-life in vitro before committing to the temporal, monetary, and ethical responsibility of constructing a mouse model to observe the physiological consequences. A suitable in vitro assay could keep everything other than the 3′UTR constant, allowing direct comparison of a gene-of-interest's 3′UTR and its proposed replacement. That in vitro work is the focus of this thesis project.

In vitro comparison of 3′UTR effects on protein expression has been done before, but with greater inconvenience. Kakoki et al. used flow cytometry of mouse embryonic stem cells, which required genomic sequence insertion and colony-picking of drug-resistant cells (Kakoki et al., 2004). Zhao et al. used lentiviruses and quantitative real-time PCR (polymerase chain reaction) (Zhao et al., 2017), which require biocontainment precautions and painstaking following of MIQE guidelines (Bustin et al., 2009). Promega Corporation's Dual-Luciferase Reporter Assay System (catalogue number E1960), instead, only requires cell line cells to undergo co-transfection, culturing, and lysis prior to measurement. In the dual-luciferase assay, Firefly luciferase and Renilla luciferase are transcribed from identical promoters with optionally different 3′UTRs (Fig. 4, 5, and 6); transcription is from two co-transfected plasmids rather than a single plasmid so that the same promoter can be used while co-transfecting in a molar ratio that avoids trans effects.

Having an internal control allows normalisation for not only transfection efficiency differences, but also any effect on transcription or translation efficiency resulting from different amounts of reporter protein generated.

In this thesis, mouse rather than human 3′UTRs have been used for the sake of deciding what genes to target in future conditional knockup in vivo mouse models. However, a human embryonic kidney cell line was used as it was already possessed by the Andressoo research group, while also having fast growth with high transfection efficiency (data not shown).

By the end of the thesis project, the dual-luciferase assay was used to compare the influence on protein expression of the native 3′UTRs of twenty mouse genes to the bGHpA flex-cassette in both its backward and excision-flipped orientations (Fig. 5). These genes were selected from the neurodegeneration-relevant categories of aggregation, anti-aggregation, and neuron survival, and include the mouse GDNF gene for which in vivo testing has previously been done.

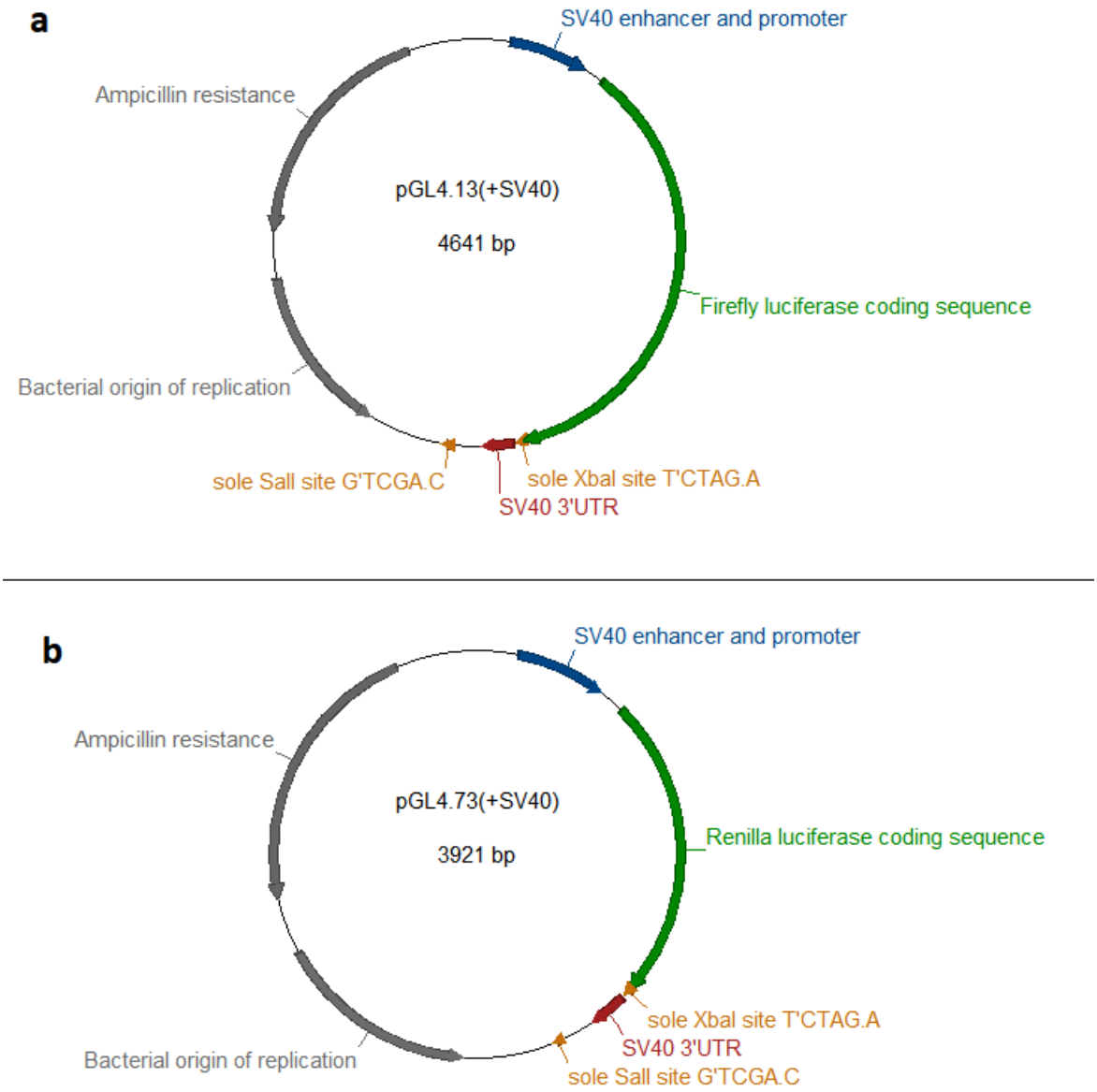

Figure 4. Plasmid structures of **(a)** pGL4.13 and **(b)** pGL4.73. The two plasmids are identical except for the luciferase coding sequence. pGL4.13 was used as an unaltered co-transfection internal control, and pGL4.73 was used as a backbone for 3′UTR modification. The luciferase is expressed eukaryotically while the Ampicillin resistance is expressed prokaryotically. The illustrations are adapted and modified from the 'Graphic Map' display option of the plasmid-map software ApE version 2.0.51. No restriction sites are displayed other than those of XbaI and SalI, which were used in this thesis for 3′UTR modification.

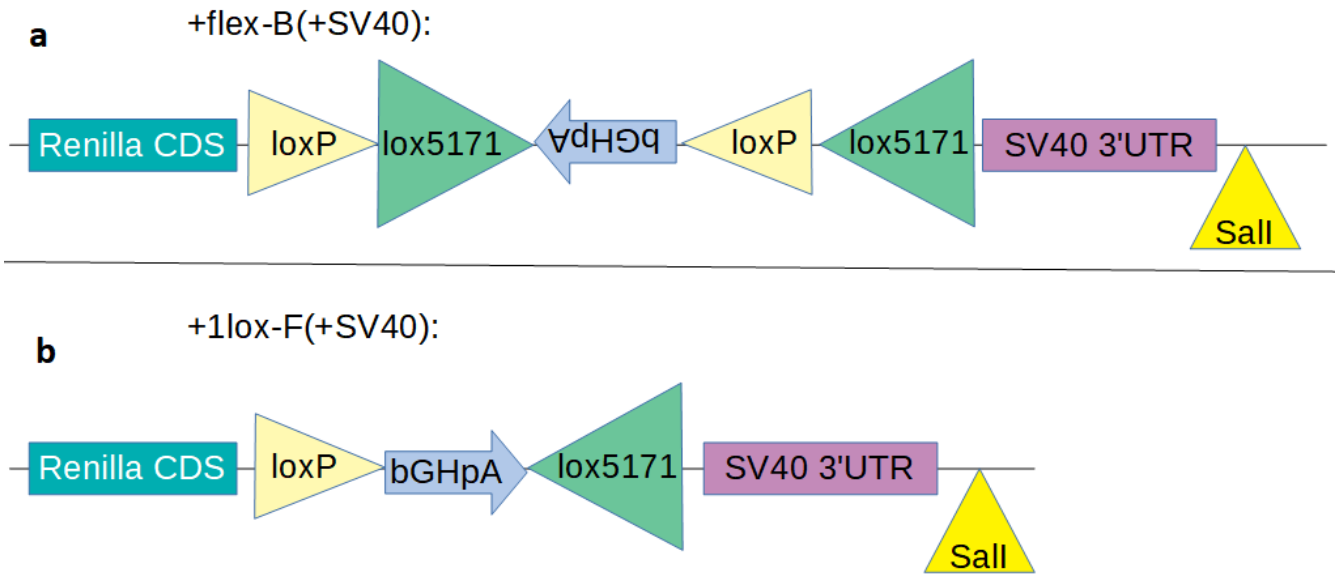

Figure 5. bGHpA flex-cassette (Schnütgen et al., 2003 and 2005) in (**a**) unflipped and (**b**) excision-flipped orientations upstream of the SV40 3′UTR in plasmid pGL4.73 (Fig. 4b). The pGL4.73 sole SalI restriction site is marked, but the pGL4.73 XbaI restriction site between the luciferase coding sequence and SV40 3′UTR no longer exists after flex-cassette insertion.

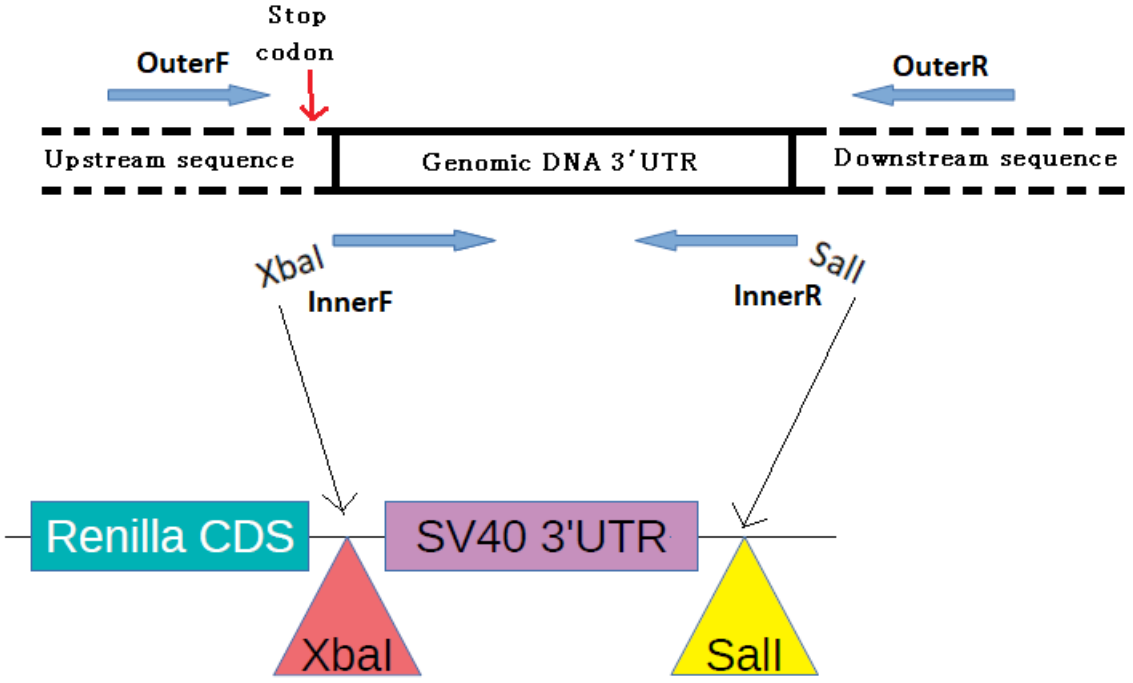

Figure 6. Genomic 3′UTR PCR primer positions, not to scale and without marked introns. Outer nested primer (**OuterF, OuterR**) and restriction-site-extension inner nested primer (**InnerF, InnerR**) annealing locations are marked with blue arrows. These primers were used for PCR of mouse genes' 3′UTRs (without including the stop codon) in order to replace the SV40 3′UTR following the Renilla luciferase coding sequence in plasmid pGL4.73 (Fig. 4b). Where the mouse gene 3′UTR contained an XbaI restriction site, an NheI or BcuI restriction site was used instead in the InnerF extension to produce an identical 5′ overhang in later restriction digestion without also cutting apart the PCRed 3′UTR.

4. Aims

3′UTR expression levels from twenty mouse genes were measured with two purposes in mind. If the expression level of a 3′UTR is much lower than that of the bGHpA flex-cassette, then that 3′UTR is a candidate for upregulation with the bGHpA flex-cassette in a mouse model. If instead the expression level of a 3′UTR is much higher than that of the bGHpA flex-cassette, then that 3′UTR's sequence is a candidate for replacement of the bGHpA sequence in the bGHpA flex-cassette, in order for the flex-cassette to be more effective in inducing expression level upregulation.

One aim of this project is thus the sorting of different genes' 3′UTRs' with these two purposes in mind.

If there were correlation between gene category and 3′UTR expression level, this was also viewed as worth noting for relevance in future research, such as whether disease-associated aggregation proteins associated consistently exhibited 3′UTRs with high expression levels.

The bGHpA flex-cassette in different orientations was measured both for assessment of the twenty mouse genes' 3′UTRs' expression levels and to compare its in vitro behaviour to in vivo behaviour previously observed. Namely, an aim was to observe whether a reverse-orientation bGHpA flex-cassette at the start of a 3′UTR had negligible effect on that 3′UTR's expression level and whether a forward-orientation bGHpA flex-cassette at the start of a 3′UTR replaced that 3′UTR's expression level with an expression level much higher than that from mouse GDNF's 3′UTR.

5. Materials and Methods

5.1. Plasmid amplification

Large yields of Promega Corporation's pGL4.13 Firefly luciferase plasmid (Fig. 4a) were produced for use as an internal control in all transfections. A large yield of Promega Corporation's pGL4.17 Renilla luciferase plasmid (Fig. 4b) was likewise produced, but this instead both for use as a plasmid backbone to prepare multiple reporter plasmids, and as an unmodified high-expression reference reporter to be transfected many times.

In all large-yield plasmid replications in this project, the plasmid underwent heat-shock transformation into calcium chloride chemically competent DH5alpha E. coli bacteria. After 1-hour outgrowth in super optimal broth with catabolite repression (SOC medium), the transformed bacteria were cultured overnight in 100 millilitres (ml) of LB medium containing 100 micrograms per millilitre (µg/ml) Ampicillin. A small portion of that culture was separated to make a −80°C bacterial glycerol stock and the rest made into a pure extracted-plasmid midiprep with Macherey-Nagel's NucleoBond Xtra Midi kit (catalogue number 740410.50). The purpose of the glycerol stock was to allow reculturing without retransformation should replenishment of a stored plasmid become necessary.

Where large yields were unnecessary, as for a reporter plasmid to be transfected a small number of times, a miniprep extraction was carried out instead from a 3 ml culture with Macherey-Nagel's NucleoSpin Plasmid miniprep kit (catalogue number 740588.250).
Throughout the thesis project, all midiprep and miniprep extractions were performed with these two kits except where specified otherwise.

The concentration of each plasmid preparation was measured with Thermo Fisher Scientific's NanoDrop 2000 Spectrophotometer (catalogue number ND-2000) following vortexing with Scientific Industries's Vortex Genie 2 (catalogue number SI-0236).

For convenience of dilution and of concentration-measurement blanking, all plasmid elution and dilution was done with sterile water rather than extraction kit elution buffer.

5.2. Reporter plasmid construction

The (+SV40) 3′UTR type plasmid (all 3′UTR abbreviations explained in Appendix A) is the unmodified pGL4.73 plasmid (Fig. 4b) containing the simian vacuolating virus 40 (SV40) 3′UTR sequence. Other reporter plasmids were constructed by inserting sequences before pGL4.73's SV40 3′UTR or by replacing the SV40 3′UTR sequence entirely (Fig. 5 and 6). The +mGDNF+60p(+SV40) 3′UTR type plasmid alone was made in the same way by an Andressoo research group former member in an earlier project.

5.2.1. Primer design

For genes-of-interest, mouse gene 3′UTR sequences were obtained initially from the National Center for Biotechnology Information (NCBI) Gene database (https://www.ncbi.nlm.nih.gov/gene/), and later from the Ensembl Genome Browser (https://www.ensembl.org/index.html), as indicated in Appendix A.

The bGHpA flex-cassette and WPRE+hGH sequences used for certain 3′UTR types were amplified from plasmids already possessed by the Andressoo research group. Appendix B records template sequences from these plasmids which were used for primer design and PCR.

PCRs from genomic DNA were initially carried out in the same way as for plasmids and Bacterial Artifical Chromosome (BAC) clones, with one primer pair each. The primers-half-annealed temperature (Tm) was also initially different for each primer pair, and thus had PCR programs with different annealing temperatures.

For reporter plasmids designed later, genomic DNA PCRs were carried out with nested primers in order to increase PCR specificity and the yield of the desired product. Outer nested primers were designed with NCBI's Primer-Blast (https://www.ncbi.nlm.nih.gov/tools/primer-blast/), while inner nested primer design was carried out in the same way as for plasmid and BAC clone PCRs.

For most genomic 3′UTR sequences used, forward and reverse inner nested primer annealing-region sequences were chosen with their 5′ ends at the beginning and end respectively of the reference sequence's annotated 3′UTR (Fig. 6). Where specified (the '-extended' 3′UTR types), the reverse primer 5' end was selected with Primer-Blast so as to include at least 500 base pairs (bp) from after the annotated end of the 3′UTR. Primer 3′ ends were initially designed according to Primer-Blast-predicted PCR product specificity, then later according to an annealing-region Tm of 60°C as predicted by the plasmid-map software ApE (‘A plasmid Editor’) version 2.0.51.

After the design of inner nested primers' annealing-regions, desired restriction sites for cloning (described further in subsection 5.2.3) were added to their 5′ ends, together with a further-five-prime AT-rich 6-nucleotide leading sequence to enable restriction digestion (NEB, ‘Cleavage Close to the End of DNA Fragments’) while mitigating any increase of non-specific annealing. Primer-dimer formation risk was predicted with Thermo Fisher Scientific's Multiple Primer Analyzer (https://www.thermofisher.com/fi/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/multiple-primer-analyzer.html), and primer length or leading sequences were changed as necessary to reduce primer-dimerisation likelihood.

Once the design process was complete, primers were ordered in the form of single-stranded DNA oligonucleotides from Metabion. All primers used in the course of this project are listed in Appendix C.

5.2.2. PCR of 3′UTR sequence inserts

To increase the yield and specificity of PCR products, Touchdown PCR (Table 1) using Thermo Fisher Scientific's Phusion High-Fidelity DNA Polymerase (catalogue number F530S) was carried out. The PCR machine used was Applied Biosystems's SimpliAmp Thermal Cycler (catalogue number A24811).

Mouse genomic DNA was obtained from mixed-background mouse tail tissue samples with VWR International's Quantabio AccuStart II Mouse Genotyping Kit (catalogue number 733-2236).

Outer nested primer PCR product samples were used as inner nested primer PCR templates, after dilution with sterile water such that the inner nested primer PCR mixture only possessed one thousandth the concentration of genomic DNA of the outer nested primer PCR mixture. This was after an agarose gel comparison of the effect of different dilutions (data not shown), and with the purpose of further reducing undesired PCR products.

Where there was persistent difficulty in amplifying specific 3′UTR sequences from genomic DNA, mouse BAC clones in E. coli strain DH10B were ordered from the BACPAC Resources Centre, cultured with 12.5 μg/ml chloramphenicol, then the BAC clones extracted with DNA precipitation and repeated 70% ethanol washing instead of with a spincolumn, as the BAC clone sizes exceeded 160 kilobases (kb) and the Macherey-Nagel miniprep kit's manual described it as only suitable for vectors smaller than 50 kb (NucleoSpin Plasmid manual Table 1). In retrospect it should have been possible to use the Macherey-Nagel midiprep kit instead, as its manual describes it as suitable for BACs and other very large constructs between 3 kb and 300 kb (NucleoBond Xtra Midi manual section 2).

Table 1. The Phusion Touchdown PCR program used by the end of the project. kbp is kilobase pairs (of DNA).

	Denaturing	**Annealing**	**Extension**	**Hold**
Stage 1 (1 cycle)	30 seconds 98°C	-	-	-
Stage 2 (10 cycles)	10 seconds 98°C	30 seconds 70-to-61°C (AutoDelta -1°C each cycle-end)	30 seconds per kbp of longest product 72°C	-
Stage 3 (20 cycles)	10 seconds 98°C	30 seconds 60°C	30 seconds per kbp of longest product 72°C	-
Stage 4 (1 cycle)	-	-	-	Until retrieval 4°C

5.2.3. Restriction digestion for cloning

3′UTR PCR products and Promega Corporation's pGL4.73 plasmid (Fig. 4b) were digested separately with Thermo Fisher Scientific's FastDigest restriction enzymes, specifically FastDigest XbaI, FastDigest NheI, FastDigest BcuI, and FastDigest SalI as required (respectively catalogue numbers FD0684, FD0973, FD1253, and FD0644). FastDigest XbaI, FastDigest NheI, and FastDigest BcuI produce identical 5′ overhangs (Table 2), and so FastDigest NheI or FastDigest BcuI were used when a 3′UTR sequence contained XbaI restriction sites. No such replacement was needed for FastDigest SalI. Digested pGL4.73 plasmids were dephosphorylated in the same reaction with Thermo Fisher Scientific's FastAP Thermosensitive Alkaline Phosphatase (catalogue number EF0651) to prevent vector-vector ligation.

Table 2. Restriction sites of the restriction enzymes used, for which XbaI, NheI, and BcuI produce identical 5′ overhangs. ' refers to a sense strand cut and . refers to an antisense strand cut.

Restriction site (FastDigest) enzyme	**Restriction site sequence**
XbaI	T ' CTAG . A
NheI	G ' CTAG . C
BcuI	A ' CTAG . T
SalI	G ' TCGA . C

5.2.4. Electrophoresis and gel purification

For the sake of a low proportion of non-specific PCR products in ligation, digested PCR products were purified from agarose gels after electrophoresis. In order to reduce undigested plasmid DNA, the digested pGL4.73 plasmid backbone used for plasmid construction was similarly purified from an agarose gel after electrophoresis.

Agarose gels were made from Tris-acetate-EDTA (TAE) buffer prepared by the University of Helsinki and agarose from Thermo Fisher Scientific (catalogue number 10776644). 1% agarose gels were used for desired product lengths over 750 bp and 3% for under 750 bp, to maximise differences in running distances (Lee et al., 2012). To prevent well overflow, agarose gel volumes were 100 ml when loading samples of 12 microlitres (µl) or less and 150 ml if loading any sample between 12 and 30 µl. Sample lane wells were prepared during gel setting using a 20-well Bio-Rad comb (catalogue number 1704448).

Initially, agarose gels were made with ethidium bromide and DNA bands visualised with ultraviolet (UV) transillumination during excision. To avoid DNA damage from using a UV transilluminator (ThermoFisher Scientific, 'SYBR Safe - DNA Gel Stain'; Gründemann and Schömig, 1996), later gels were instead made containing 0.001% crystal violet (from Thermo Fisher Scientific, catalogue number 10637884), for which high-yield DNA bands are visible in normal light conditions (Rand, 1996).

In the final workflow, to allow visibility of low-yield DNA bands while preventing UV DNA damage, agarose gels were made with 0.0001% SYBR Safe DNA Gel Stain (from Invitrogen, catalogue number S33102), and gel slices containing DNA bands were excised using Bio-Rad's ChemiDoc MP Imaging System (catalogue number 12003154) and its Blue Sample Tray (catalogue number 12003027), together with Bio-Rad's XcitaBlue Viewing Goggles (catalogue number 1708185) for filtering out blue light while leaving the emitted fluorescence visible.

Thermo Fisher Scientific's GeneRuler 1 kb DNA Ladder (catalogue number SM0311) was used in all gels to identify approximate DNA molecule lengths.

Agarose gel electrophoresis itself was carried out in Bio-Rad's Sub-Cell GT Horizontal Electrophoresis System, 15x25 cm tray (catalogue number 1704404), for 3 hours at 55 volts (V). The purification kit used after gel slice excision was Macherey-Nagel's NucleoSpin Gel and PCR Clean-up kit (catalogue number 740609.250), and the running voltage below 60 V was selected according to its manual's recommendation (NucleoSpin Gel and PCR Clean-up manual subsection 2.5). Following purification, DNA concentration was measured with with the same NanoDrop spectrophotometer used for plasmid concentration measurement. As with plasmid extraction, sterile water was used in place of elution buffer for elution and dilution.

5.2.5. Ligation of insert and backbone

Ligation of purified inserts into purified plasmid backbones were carried out with Thermo Fisher Scientific's T4 DNA Ligase (catalogue number EL0014).

After transformation (as in subsection 5.1), colonies were grown on LB agar plates containing 100 μg/ml Ampicillin. Liquid cultures were grown from individual colony samples, then miniprep plasmid extractions carried out (also as in subsection 5.1). Restriction analysis was done with 30-minute 120 V ethidium bromide agarose gel electrophoresis and UV transillumination for initial analysis of the clones by comparing observed DNA fragment sizes to fragment sizes predicted by ApE version 2.0.51. For seemingly-successful plasmid constructs this was followed by Sanger sequencing, more expensive than restriction analysis, for final confirmation. Sanger sequencing was also carried out to check the +mGDNF+60p(+SV40) 3′UTR type plasmid previously produced by an Andressoo research group former member.

5.3. Co-transfection

Using Thermo Fisher Scientific's TurboFect transfection reagent (catalogue number R0531), the HEK-293 human embryonic kidney cell line (ATCC number CRL-1573) was co-transfected in a 96-well cell culture plate's well with a subsection 5.2 reporter plasmid together with Promega Corporation's pGL4.13 plasmid (Fig. 4a) as an internal control. Co-transfection masses were calculated according to the NanoDrop-measured DNA concentrations and predicted reporter plasmid lengths to implement a consistent molar ratio of internal control plasmid to reporter plasmid within wells measured on the same day.

All wells transfected in the same day were also measured in the same day. Any well transfected was part of a replicate of three wells transfected with the same co-transfection mix. Each cell culture plate used for transfection contained replicates for different reporter plasmids which could thus be directly compared to each other. Identical co-transfection plans were carried out on three different days where time allowed, each plate's wells to undergo measurement together on a different day.

5.4. Measurement recording

Firefly and Renilla luminescence was induced in turn with the luciferase substrates of Promega Corporation's Dual-Luciferase Reporter Assay System (catalogue number E1960) and measured with Promega Corporation's GloMax 20/20 Luminometer (catalogue number E5311), which measures luminescence from one Eppendorf tube at a time and thus avoids all crosstalk. The measurement program was a 2-second preread delay followed by a 10-second measurement period, first after mixing with the Firefly luciferase substrate and again after then mixing with the Renilla luciferase substrate, which also quenches Firefly luciferase activity. As the measurement steps should be done quickly and identically, cell culture plate well lysates were all transferred to labelled Eppendorf tubes after cell lysis to prevent losing track of wells while making measurements.

Normalised measurements for the wells were calculated as Renilla luminescence divided by Firefly luminescence, to correct for differences in cell number and expression efficiency between wells. As the GloMax 20/20's detector is a photomultiplier tube (Glomax 20/20 manual subsection 12.G.) which gives lognormally-distributed results (Kissick et al., 2010), measurements can be logarithmised to produce normally-distributed values. Unlike lognormal distributions, normal distributions ideally have symmetrical distribution about a true parameter being measured, and thus are more visually intuitive for estimating that true parameter. Logarithmisation to the base 2 was chosen, with a distance of 1 equating to a 2-fold difference. As log2(x/y) = log2(x) - log2(y), and the difference of normally-distributed random variables is also normally-distributed, each Renilla/Firefly normalised measurement in each replicate was logarithmised directly to produce normally-distributed data points.

5.5. Statistical analysis

5.5.1. Testing of samples for homogeneity of variance

Using logarithmised measurement ratios, Levene's test was carried out with PSPP version 1.2.0. Empirical cumulative distribution function comparison to chi-squared distributions was carried out using the CHISQDIST function of LibreOffice Calc in LibreOffice version 6.2.8.2 (x64).

5.5.2. Testing of samples for normality

Sample residuals underwent normality testing in R version 3.6.2. A Shapiro-Wilk normality test was carried out with shapiro.test(), and a Q-Q plot for normality testing was carried out with qqnorm() and qqline().

5.6. Day-adjustment normalisation of recorded data points

As suggested by the dual-luciferase assay's manual, the freezing and thawing history of the Firefly-luciferase's substrate can change the amount of light produced per luciferase molecule (Dual-Luciferase Reporter Assay System manual section 5 and subsection 6.A.). These multiplicative changes to the luminescence ratios are presented as common additive offsets in the logarithmised values, and thus weighted averages of common 3′UTR types in different measurement days were used to normalise for differences between those days.

Further, while measurement days 1-0, 1-1, and 1-2 (measurement day abbreviations explained in Appendix A) had 1:11.8 control:reporter co-transfection molar ratios (consisting of a constant 10 nanograms (ng) internal control plasmid co-transfected with a reporter plasmid mass which was length-proportionate to 100 ng of unaltered pGL4.73), measurement days after those used transfections done with a 1:10 co-transfection molar ratio with a constant 200 ng total co-transfection mass, as recommended by the TurboFect ‘Transfection Reagent Considerations’ user guide (Thermo Scientific TurboFect Transfection Reagent protocol Table 1). Co-transfection ratio differences between measurement days should give rise to multiplicative differences which can similarly be corrected for through logarithmised-value additive-offset day-adjustment.

Though (+SV40) was the nominal reference 3′UTR type, day-adjustment was based on all common 3′UTR types instead of (+SV40) alone so as to lower the influence that any single replicate's variation had on the adjustment process. Specifically, each common 3′UTR type contributed to each day's day-adjustment offset according to that 3′UTR type's widest confidence interval in all days being compared.

When multiple days were normalised relative to each other, to calculate relative additive offsets the same set of weights across days was used to make a weighted average of all arithmetic means of all common 3′UTR type samples for each day, those values being logarithmised measurement ratios as described above. Sample means were used as estimates of the ‘true’ value represented by the normally-distributed replicate measurements. Each weight was the minimum inverse-variance of that 3′UTR type's relative expression level estimate within those days, so that a 3′UTR type's estimate's low variance in one day could not impart that 3′UTR type high weighting for a day on which its estimate had high variance.

After day-adjustment, a single common offset was used to set the approximate expression of the nominal reference 3′UTR type (+SV40) to '0' (representing 100% expression) for more intuitive evaluation of other 3′UTR types as relative to it.

5.7. Graphical representation of data points

Data points were charted as strip-chart scatterplots in R version 3.6.2, allowing display of all values without needing to calculate a single estimate for each 3′UTR type. The R package ‘vipor’ (VIolin POints in R) (RDocumentation, ‘vipor package’) was used to implement quasirandom jitter so that nearby points were visually distinct without overlapping.

6. Results and Discussion

6.1. Differing plasmid DNA degradation and the implementation of ‘batches’

For the first three measurement days (1-0, 1-1, 1-2), co-transfection plasmid mixtures were mixed on each transfection day (one transfection day per measurement day). This was changed after failed measurement days in which Firefly luciferase expression measurements were unusably low, unaffected by the assay substrate's condition (data not shown). Upon testing, it was revealed that plasmids stored in different tubes were degrading at different rates, due to shearing when vortexing (Wu et al., 2009, Fig. 8) or storage conditions or otherwise, and in doing so displaying effective transfection efficiencies which were not representative of their measured concentrations (Fig. 7).

As NanoDrop concentration measurement cannot distinguish between degraded and undegraded DNA, one option considered was to make fresh plasmid preparations for each measurement day, exposing each plasmid type to the minimum possible time and vortexing before transfection. To save time, a ‘batches’ approach was instead adopted in which groups of reporter plasmids and their internal control plasmid were prepared together, co-transfection mixtures made soon after plasmid preparation, and those same mixtures used in all measurement days for that batch. The intention was that, even if different co-transfection mixtures underwent different degrees of degradation, the reporter and internal control plasmids in each mixture would undergo the same degree of degradation and maintain the same relative effective transfection efficiency.

As inaccuracies in both NanoDrop concentration measurement and human pipetting can introduce ratio-variation stochastic error, this approach of using the same mixture for multiple measurement days unfortunately limits the number of ratio-variation instances for each 3′UTR type to the number of batches in which that 3′UTR type is measured, rather than every measurement day having a different error to contribute to averaged noise. However, day-adjustment normalisation of measurement-day substrate-efficiency differences within batches becomes more reliable.

When using batch grouping, data point day-adjustment (subsection 5.6) between different batches requires those batches to share at least one 3′UTR type, rather than staggering 3′UTR type measurement as new types are added and thoroughly-tested types retired. However, this can be bypassed through two measurement days for different batches being carried out in the same session with the same luciferase substrates and same co-transfection molar ratio, in which case day-adjustment is unnecessary as the day is already the same. This was done for measurement day 2-2/3-1 to anchor batches 2 and 3 to each other, though there was no similar double measurement day to anchor batch 1 to batch 2 or 3.

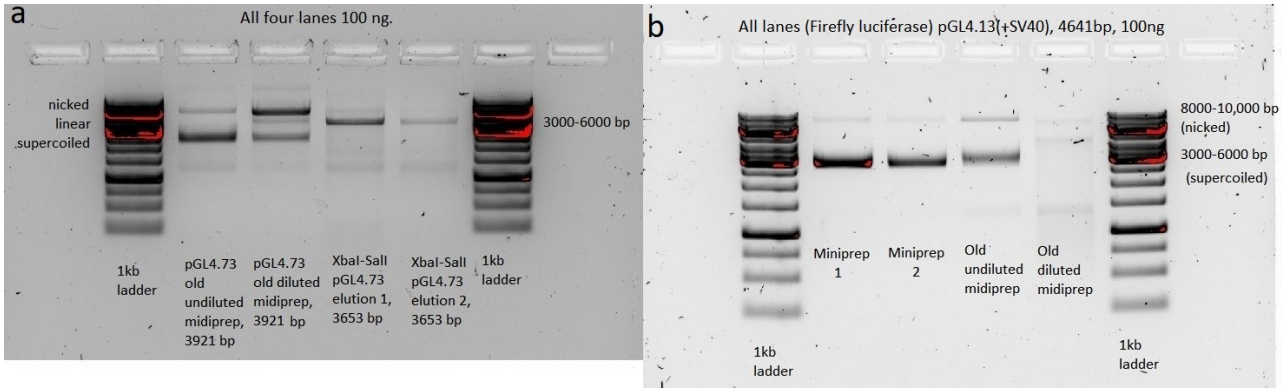

Figure 7. DNA degradation of plasmids used in co-transfection, suspected to be due to vortex mixing. 100 ng DNA masses according to NanoDrop concentration measurements were subjected to electrophoresis, then the DNA band intensities of different plasmid isoforms (supercoiled, nicked, and linear (Tirabassi, n.d.; Oppenheim, 1981)) observed. The DNA ladder is Thermo Fisher Scientific's GeneRuler 1 kb DNA Ladder.

(a)

Sample lane 1 contains rarely-vortex-mixed undiluted purified pGL4.73 plasmid DNA (Fig. 4b). It displays a strong supercoiled-plasmid band, a weak nicked-plasmid band, and no visible linear-plasmid band.

Sample lane 2 contains pGL4.73 plasmid originally diluted from sample lane 1's source tube, frequently vortex-mixed for use in co-transfections. It displays a strong nicked-plasmid band, a medium-strength supercoiled band, and a weak linear-plasmid band.

Sample lane 3 contains linear DNA derived from digesting plasmid DNA from sample lane 1's source tube with FastDigest XbaI and FastDigest SalI, then purifying with a NucleoSpin spincolumn.

Sample lane 4 contains DNA eluted in a second elution from the spincolumn of sample lane 3. Sample lanes 3 and 4 together show how the NanoDrop-measured DNA concentration of a second-elution eluate is untrustworthy, thought to be due to a higher chaotropic salt concentration relative to DNA concentration, as reflected in a NanoDrop absorption ‘peak’ more resembling a diagonal line than a peak (data not shown).

(b) All samples are pGL4.13 plasmid DNA (Fig. 4a).

Sample lanes 1 and 2 contain freshly-prepared plasmid from two different minipreps. It shows strong supercoiled-plasmid bands and weak nicked-plasmid bands. Despite the NanoDrop-measured ostensibly equal masses, the two samples' band strengths are visibly different, thought to be due to different chaotropic salt proportions (as in (a) sample lane 4).

Sample lane 3 contains rarely-vortex-mixed undiluted purified pGL4.13 plasmid, analogous to (a) sample lane 1. It displays a weaker supercoiled-plasmid band and stronger nicked-plasmid band than samples lanes 1 or 2.

Sample lane 4 contains pGL4.13 plasmid originally diluted from sample lane 3's DNA and frequently vortex-mixed for use in co-transfections, analogous to (a) sample lane 2. It displays weak nicked- and linear-plasmid bands and no visible supercoiled band.

6.2. Statistical analysis

Readers wishing to carry out alternative statistical analysis can find all measurement data in Appendix D.

6.2.1. Initial intentions

Compared to Genetics, my Statistics foundation is much shallower. It was initially imagined, being able to obtain an expression level estimate (mean) and confidence interval (Barde and Barde, 2012) for each measurements-replicate sample, that it would be possible to combine normalised measurements under different conditions to arrive at a single estimate and confidence interval for the true expression level which took into account all sample estimates and confidence intervals. At the time, it was imagined that resultant estimates would have smaller confidence intervals for consistent contributing estimates (such as estimates within each other's confidence intervals) and wider confidence intervals for more divergent contributing estimates (such as estimates falling outside each other's confidence intervals).

The intention was to use 3-sigma (99.7%) confidence intervals, as a 95% confidence interval would be expected to exclude an estimated true parameter 1 in 20 times (compared to 3 in 1000 times), and this thesis project uses measurements from more than 20 3′UTR types.

In practice, first would have been obtained a single estimate and confidence interval per 3′UTR type per measurement day, then per 3′UTR type per batch, and then finally only per 3′UTR type.

It was then planned, upon obtaining a single estimate and confidence interval for every 3′UTR type's relative expression level, to learn and carry out an appropriate statistical analysis method for numerically estimating how distant different parameters were likely to be from each other. Considerations planned to make sure to be taken into account or else explicitly acknowledged as problems were the unreliability of p-values (Cohen, 1994; Greenland et al., 2016) and the multiple comparisons problem. For the unreliability of p-values there was curiosity about whether a Bayesian method could be used to directly report subjective likelihood ratios (Cohen, 1994; Kass and Raftery, 1995). For the multiple comparisons problem the term ANOVA (ANalysis Of VAriance) had been previously encountered, said to compare all estimates to each other in a single operation.

These plans collapsed upon being unable to obtain a meaningful way of combining different estimates.

A sample mean represents two forms of bias, namely the distance of the sample mean from the sample's population mean and the distance of the sample's population mean from the true value to be measured. The first bias originates from different measurement conditions within a replicate (such as slightly different delays when operating measuring equipment); the second bias originates from different measurement conditions between replicates (such as slightly different mixing ratios when mixing reporter and internal control plasmids for co-transfection). Systematic contributors to bias can be partly corrected for by normalisation, while stochastic contributors can be partly corrected for by averaging. When a sample has a small confidence interval, it represents that the within-replicate bias is low.

Notably, for normally-distributed values such as that this thesis's logarithmised measurement ratios were intended to be, a sample's mean (and its distance from the sample's population mean) and its sample variance are independent (Geary, 1936; Knight, 2000, proposition 2.11)--the confidence interval is dependent only on the sample size and the sample's population variance. However, the estimated population variance is still calculated from the sample variance, whether from averaging infinite sample variances which have undergone Bessel's correction to give unbiased estimates of a common population variance, or instead using a different correction for fewer sample variances (Eq. 1).

Equation 1. Sample variance correction to use when averaging corrected sample variances of a normal distribution to estimate a common population variance, such that the mean square error of the estimate is minimised. n is the number of measurements per sample ('sample size') and m is the number of samples. For infinite samples, this approaches Bessel's correction n/(n-1), giving an unbiased-but-loose estimator, the estimates of which can be averaged to an exact value. For a single sample (in which case averaging is not possible), this is n/(n+1), giving a tight-but-biased shrinkage estimator (Singh and Saxena, 2001, Eq. 1.2 and 1.7). For 116 samples of sample size 3, this is 3/(3 - 1 + (2/116)), equal to 58/39. For 0 samples, this is undefined.

$$\text{Sample variance correction} = \frac{n}{n-1+\frac{2}{m}}$$

When calculating a final estimate from multiple sample estimates, weighting the sample estimates according to sample confidence interval decreases within-replicate bias while skewing between-replicate bias, and using no weighting gives an unskewed between-replicate bias while letting low-confidence samples introduce more within-replicate bias. I do not currently know how to calculate what weighting minimises total bias, and have no leads on how to calculate the confidence interval of a final estimate when the contributing estimates have different confidence intervals.

At this point, the ANOVA method appeared the most ready tool to hand, being a frequently-used tool to handle data with different sources of variance--in this case, the between-replicate-measurements variance, the between-measurement-days variance, and the between-batches variance. However, ANOVA requires homogeneity of variance (homoscedasticity) as an assumption. Predicting that the samples of logarithmised measurement-ratios, obtained through the same measurement procedure, would have the same variance (due to the within-replicate noise being of the same source in each case), it was thus sought to test the samples for homogeneity of variance, which if satisfactory would also allow calculation of samples' estimates' confidence intervals from estimated population variance and sample size alone, rather than depending on individual sample variances which were independent of their sample mean's distance from the population mean.

6.2.2. Insufficient evidence for homogeneity of samples' population variance

A chi-squared distribution scaled according to its degrees of freedom has a mean of 1, being the population variance of the standard normal distribution that gives rise to it. If further scaled relative to a different number, it will instead appear as though it originated from a normal distribution with a population variance of that number. Bessel-corrected sample variances of a constant sample size likewise take the form of a scaled chi-squared distribution (Knight, 2000, proposition 2.11) with an unbiased mean at the origin normal distribution's population variance, with one fewer degree of freedom than the sample size.

The 116 sample variances of sample size 3 were sorted and Bessel-corrected to form an empirical cumulative distribution function (CDF) and compared to the CDF of an ideal chi-squared distribution with 2 degrees of freedom, scaled according to the estimated population variance (~0.01192, equivalent to a standard deviation of ~1.27-fold when represented as an unlogarithmised relative expression level; 3-sigma would be equivalent to ~2.05-fold) (Fig. 8). However, the empirical CDF curve did not match the predicted scaled CDF, and instead more closely resembled a predicted chi-squared distribution with only 1 degree of freedom.

Levene's test was also carried out, understood to be a typical homogeneity of variance test for before allowing an ANOVA, on all 123 samples (116 of sample size 3 and 7 of sample size 2). The result was displayed as '.000', indicating to reject the null hypothesis of homogeneity of variance, consistent with the visual impression of the empirical CDF. As further examples, the Levene's test significances for all (+SV40) samples, all Batch 2 samples, and all Batch 2 Day 1 samples were displayed as '.006', '.000', and '.001' respectively.

While Welch's ANOVA is an ANOVA that does not require homogeneity of variance, to my current knowledge it only exists as a One-Way version and not in the Three-Way form appropriate for analysis in this instance (to account for the contributing factors of 3′UTR type, measurement day, and batch).

As the samples could not be indicated to possess a common population variance, the subsection 5.6 day-adjustment normalisation's inverse-variance weighting was carried out with population variance estimates as estimated from each sample separately (Eq. 2).

Aside from the mean square error, other estimators exist such as mean absolute error and maximum likelihood estimation which may have different advantages. Minimisation of the mean square error was chosen with the intention that by definition this would minimise the variance, and thus the confidence interval, as though taken respective to the true parameter being estimated (Murphy, 2012, subsection 6.4.4).

Equation 2. Equation for each sample's estimate's inverse-variance. One measurement in a sample is X_i, and the number of measurements in a sample ('sample size') is n. As the measurements in each sample were planned to be normally-distributed, the 'true expression level' estimate is the mean of the sample measurements, for which the variance is the sample's population variance divided by the sample size (Wellmer, 1998; Murphy, 2012, Eq. 6.43). As an assumption of common population variance could not be satisfactorily supported, each sample's population variance estimate uses the normal distribution sample variance correction for one sample only (Eq. 1).

$$\text{estimate's inverse-variance} \sim= \frac{\text{sample size}}{\text{population variance estimate}} = \frac{n}{\frac{1}{n+1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}$$

Lacking time in the remainder of the Master's degree duration for further statistics self-study, and at a loss as to how meaningful analysis of my results could be carried out, consultation was sought with University Lecturers Pekka Heino and Janne Ravantti. The conclusion was to display the results in strip-chart scatterplot form, so as to include all data in a visually-intuitive way without carrying out numerical hypothesis-testing. Confidence intervals or box plots were not used as the uncombined markings for each different 3- or 2-measurement sample for a 3′UTR type would have rendered the charts unreadable.

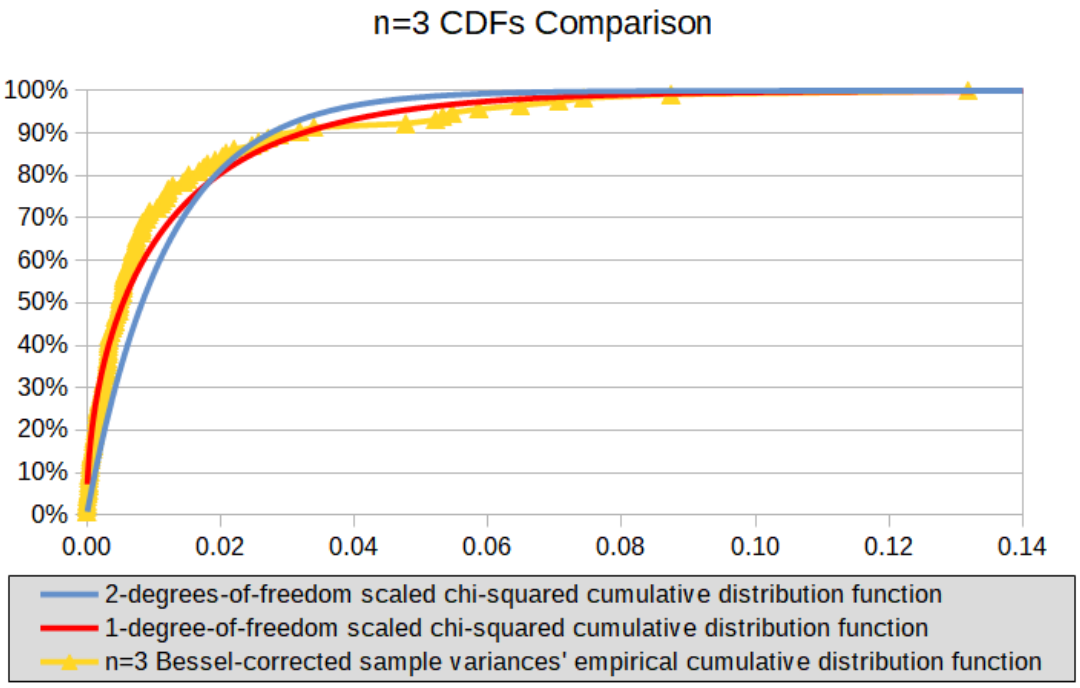

Figure 8. Curve-comparison homogeneity of variance check in LibreOffice Calc. 116 samples all of sample size 3 were used. n is sample size. Bessel-correction is multiplication of a sample variance by n/(n-1) to obtain an unbiased estimate of population variance. Sorted-and-scaled Bessel-corrected sample variances as an empirical cumulative distribution function are compared to 1-degree-of-freedom and 2-degrees-of-freedom scaled chi-squared distribution curves, charted in the form {f(x)}=CHISQDIST(1*{x}/0.01192,1,1) and {f(x)}=CHISQDIST(2*{x}/0.01192,2,1) respectively. The expectation was held that the samples' curve should most closely match the 2-degrees-of-freedom curve (Knight, 2000, proposition 2.11) if the samples possess a common population variance (estimated as 0.01192); however, it does not.

6.2.3. Insufficient evidence for samples being normally-distributed

Despite having intended that the logarithmised measurement ratios be normally-distributed, the inability to assume a common population variance casts doubt on this.

Residuals for all 123 samples were calculated by subtracting each sample's mean from its measurements, then in R a Shapiro-Wilk normality test carried out and a Q-Q plot drawn (Fig. 9). The Shapiro-Wilk result was displayed as '0.000000005978', indicating to reject the null hypothesis of a normal distribution, and the Q-Q plot appears normal in the middle while giving an impression of being too light-tailed, as of a Student's-t distribution.

The approximate symmetry of the samples' distribution bodes well for using the samples' means as expression value estimates, but the distribution not being normal bodes ill for methods using population variances estimated assuming normal distributions. Unfortunately, I currently lack methods for making estimations based on non-normal distributions.

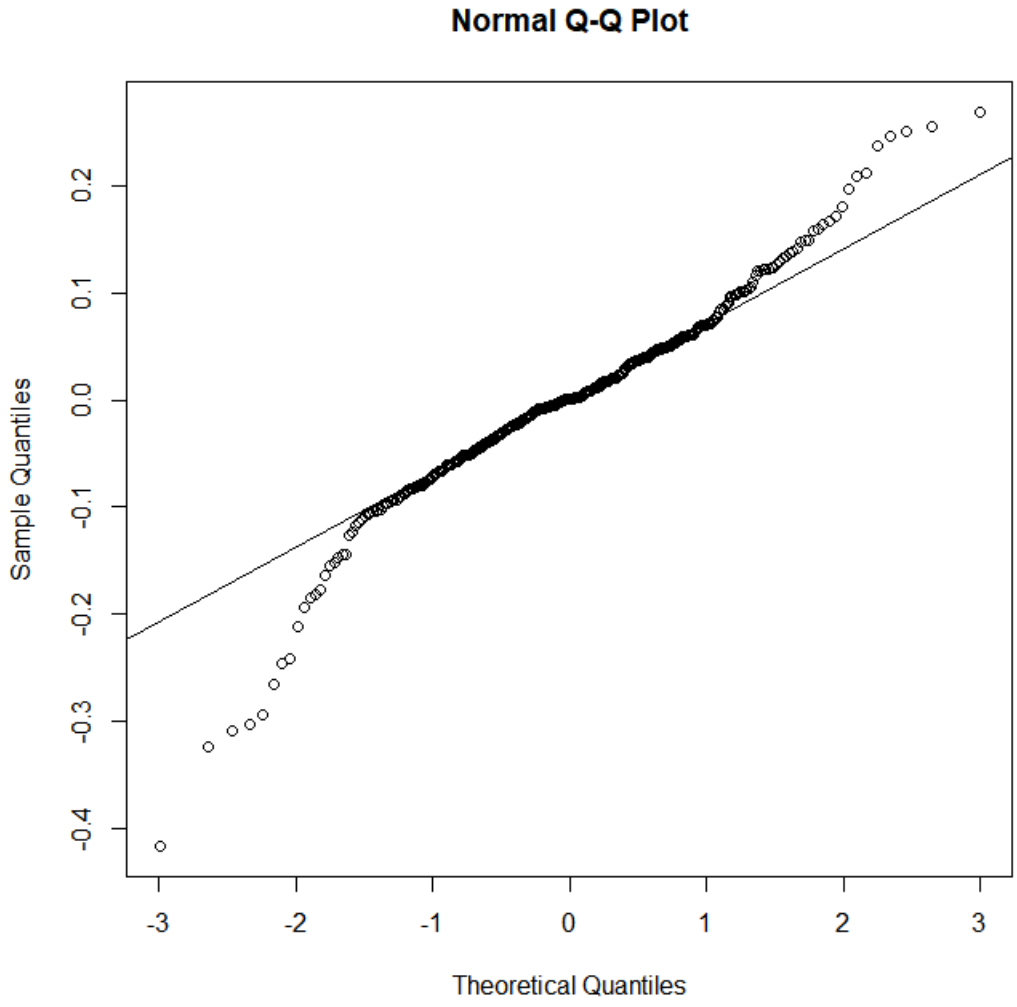

Figure 9. Q-Q plot of sample residuals for normality testing, produced in R with qqnorm() and qqline().

6.3. All measurements in unadjusted and day-adjusted forms

Figure 10 displays the data points for all 3′UTR types tested in this thesis. All raw and processed measurement values are recorded in Appendix D.

In the unadjusted data (Fig. 10a), a common offset can at a glance be observed between measurement days 2-1 and 2-2 (purple and orange circles), and between 3-1 and 3-2 (orange and green triangles). After day-adjustment (Fig. 10b), these and other data points are instead clustered closely together, assisting comparison of relative protein expression. Replicates' data points more distanced from others of the same 3′UTR type are overwhelmingly from measurement days 1-0, 1-1, and 1-2, reflecting the less reliable nature of their different workflow (subsection 6.1).

3′UTR types (+SV40), +0lox-F(+SV40), and +mDNAJB6-1 are the only ones present in both Preparation Batch 2 (circles) and Preparation Batch 3 (triangles), and the closeness of their unadjusted points on Measurement Day 2-2/3-1 (orange circles and triangles) supports that no day-adjustment is necessary when the measurement session, substrates, and co-transfection molar ratio are all the same.

In addition to flex-cassette and viral sequence testing (described in later subsections), 3′UTRs from twenty genes for anti-aggregation, aggregation (Chiti and Dobson, 2017), and neuron-survival proteins were tested both for evaluation as candidates for conditional knockup and to see whether any common aspect emerged that could shed light on neurodegeneration and possible treatment approaches. However, all three groups exhibited expressions within the same overall range with no major distinction between observed expression levels.

One noteworthy 3′UTR type was +mGDF11, which showed far lower expression than all other 3′UTR sequences tested. Upon checking, it was found that the NCBI reference sequence NM_010272.1 used for +mGDF11's primer design does not contain the expected polyadenylation signal, unlike the equivalent-at-the-time Ensembl Genome Browser reference sequence ENSMUST00000026408.6 which has a 3′UTR length of 2811 bp rather than 560 bp. At this time of writing, NCBI displays the reference sequence NM_010272.2, which has a 3′UTR length of 4582 bp. +mGDF11's low observed expression is thus predicted to be an artefact of its 3′UTR sequence being incomplete.

Most figures in later subsections show subset data from Figure 10b.

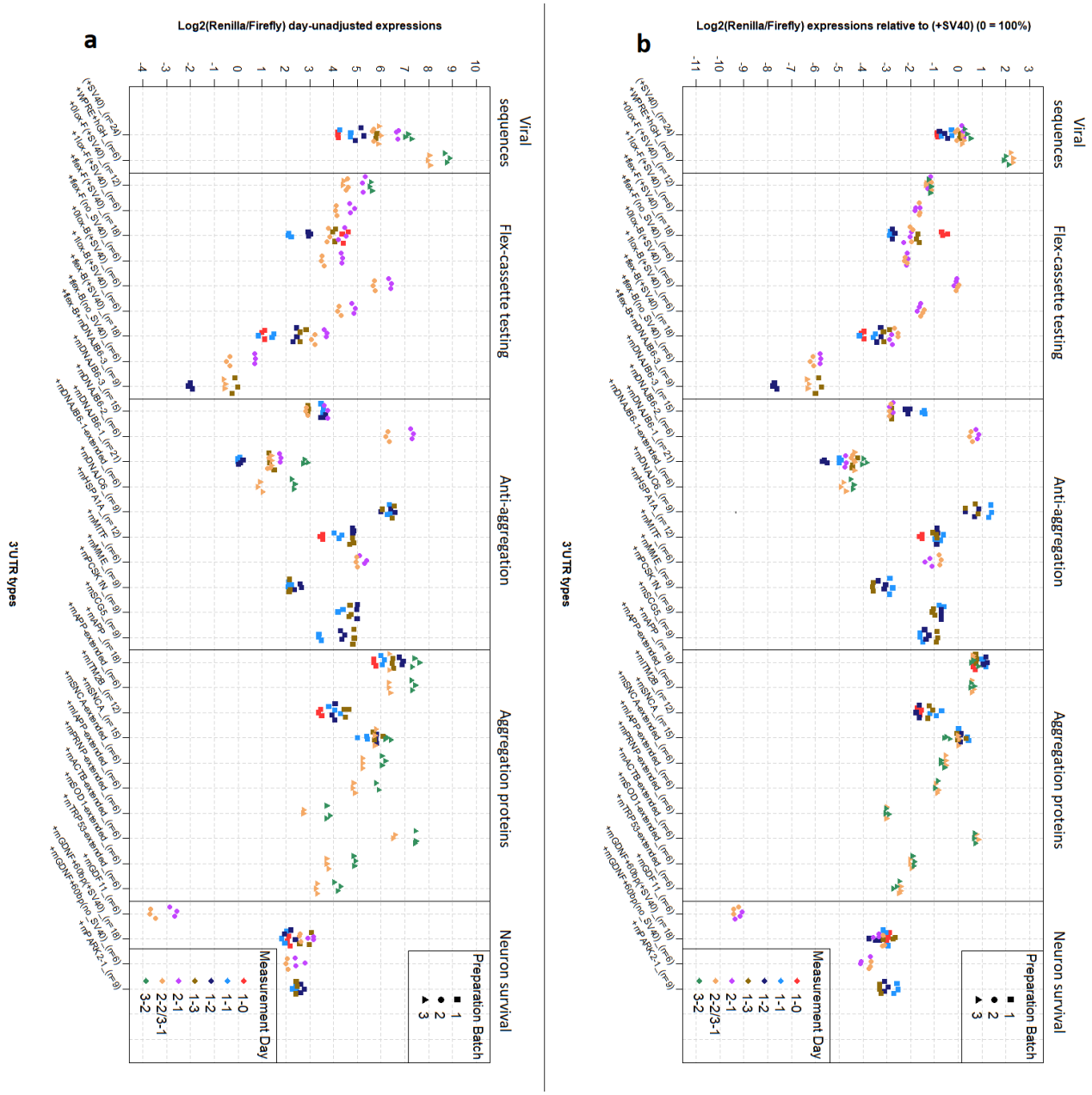

Figure 10. Full results, broken down into specific comparisons in later figures. These are the measured effects on protein expression in HEK-293 cells of all 3′UTR types described in this thesis, grouped by category. A relative expression of 0 is 100%, 1 is 200%, 3 is 800%, -1 is 50%, et cetera. Preparation Batch indicates the batch of plasmids prepared and measured at the same time, and Measurement Day is the abbreviation for each day in {Batch}-{Batch's day} format, with measurement sessions 2-2 and 3-1 performed on the same day. Total number of measurements/data points for each 3′UTR type is indicated as the suffix '_(n={measurement number})'. **(a)** Raw-data logarithmised measurement ratios with no adjustment. **(b)** Logarithmised measurement ratios adjusted for measurement day as described in subsection 5.6. All displayed data points are recorded in Appendix D, Table 7. All measurement day and 3′UTR type abbreviations are listed and described in Appendix A.

6.4. Flex-cassette conditional knockup testing

6.4.1. Initial test

The one available example of in vivo flex-cassette conditional knockup was mouse GDNF (Fig. 3). An ideal in vitro comparison to emulate this would have been a '+flex-B+mGDNF' and a '+1lox-F+mGDNF' 3′UTR types to reproduce the excision-flipping of the flex-cassette, but in the face of PCR difficulties these were not successfully constructed within the available timeframe.

Initially, +mGDNF+60bp(+SV40) and (+SV40) were compared with +flex-F(+SV40) and +flex-B(+SV40), and +1lox-F(+SV40) was constructed later together with other plasmids for more in-depth investigation. However, here the +1lox-F(+SV40) comparisons results can be shown first.

In Figure 11, the +mGDNF+60bp(+SV40) +1lox-F(+SV40) are examined to compare the expression from the GDNF 3′UTR to that of the excision-flipped (forward-orientation) bGHpA flex-cassette, if the sequence downstream of the flex-cassette is consistently cleaved away in cleavage-and-polyadenylation. The effect of an upstream unflipped (backward-orientation) bGHpA flex-cassette on the expressions of the (+SV40) and +mDNAJB6-3 3′UTR types is similarly displayed.

The forward-orientation (‘-F’) GDNF-comparison in vitro result (+1lox-F expression between 4- and 8-fold greater than for +mGDNF+60bp(+SV40)) is consistent with the in vivo result (heterozygous expression with Cre recombinase between 4- and 8-fold greater than for wild-type), keeping in mind that with binary-logarithmised values a distance of 1 equates to a 2-fold difference in expression.

By contrast, the backward-orientation (‘-B’) in vitro result (an approximate 8-fold drop in expression for both tested 3′UTR types) is wholly inconsistent with the in vivo result (heterozygous expression without Cre recombinase no lower than for wild-type).

The inability to fully reproduce the in vivo behaviour in this in vitro context warns of potential unreliability if wanting to use the same method for other testing. To learn whether the bGHpA sequence or lox sites were responsible for this, more in-depth investigation was carried out.

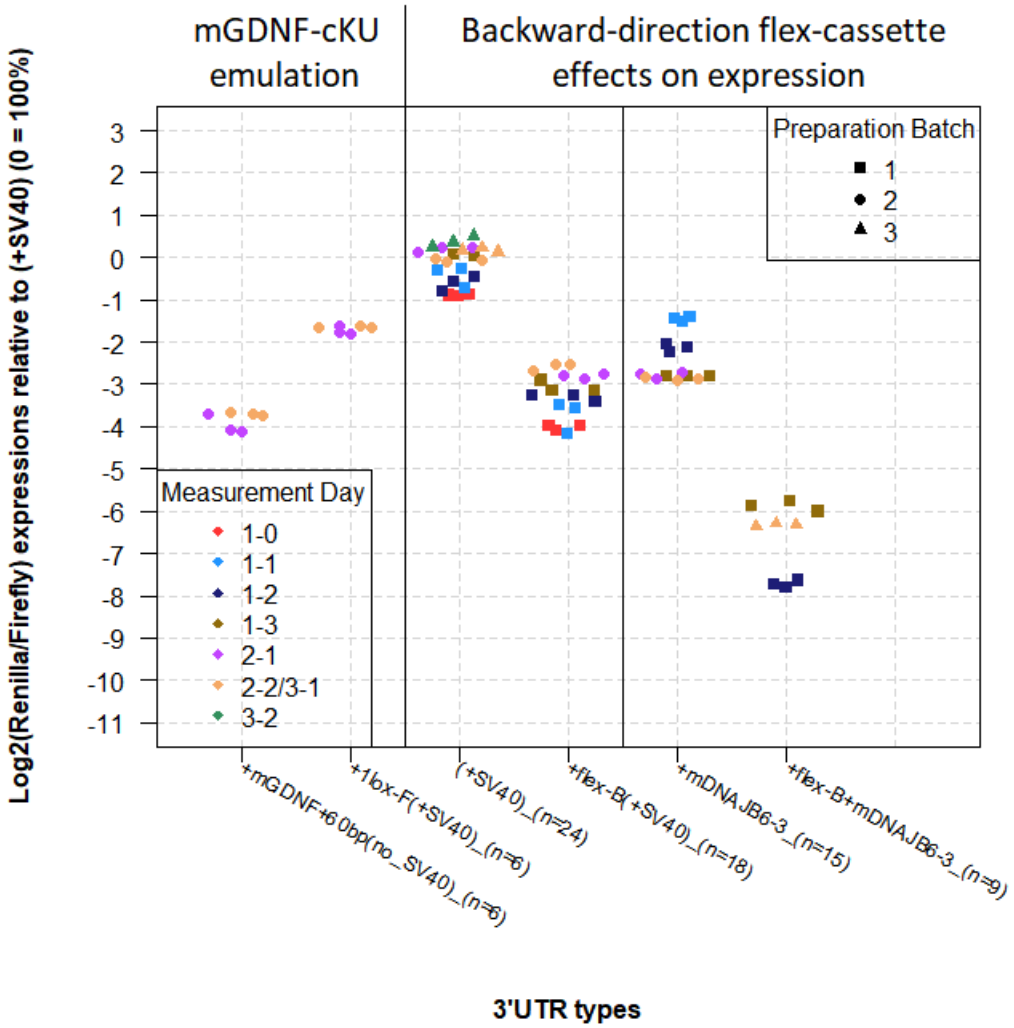

Figure 11. Initial conditional knockup text. These are the effects of an inserted upstream excision-flipped flex-cassette ('1lox-F') or unflipped flex-cassette ('flex-B') on the 3′UTR following. In this figure (unlike in Figure 12) the +1lox-F(+SV40) is only valid if replacement of the downstream 3′UTR is assumed, as a '+1lox-F+mGDNF' 3′UTR type could not be constructed in the necessary timeframe. All measurement day and 3′UTR type abbreviations are listed and described in Appendix A.

6.4.2. Readthrough-influence test

To investigate the bGHpA flex-cassette behaviour in vitro, the first underlying assumptions checked were those about its respective behaviour with the bGHpA sequence in forward and backward orientation. Namely, keeping the flanking lox sites identical for constant influence, does the forward-orientation flex-cassette reliably replace the expression of a downstream 3′UTR as intended, or is its cleavage-and-polyadenylation behaviour instead weak rather than strong (Chao et al., 1999)? Does the backward-orientation flex-cassette have zero cleavage-and-polyadenylation behaviour as intended, or does it instead replace the downstream 3′UTR? In a readthrough-influence test (Fig. 12), the downstream SV40 3′UTR was removed from +flex-F(+SV40) and +flex-B(+SV40) to test whether its influence was as intended negligible for the first and predominant for the second. The SV40 3′UTR was likewise removed from +mGDNF+60bp(+SV40) to test whether it was affecting measured expression, possible if the GDNF 3′UTR cleavage-and-polyadenylation behaviour were weak rather than strong.

Separately, most inserted 3′UTR sequences taken from mouse genes were PCRed starting from immediately after a stop-codon and ending at the point corresponding an mRNA transcript's cleavage-and-polyadenylation site: however, in mRNA transcription a longer transcript is first transcribed, then cleaved at the cleavage-and-polyadenylation site. The cleavage and polyadenylation specificity factor (CPSF) and cleavage stimulation factor (CstF) protein complexes cooperate to do this, of which CPSF binds upstream at the AAUAAA-like polyadenylation signal while CstF binds to downstream GU/U-rich sequence elements. During the binding-and-cleavage process, approximately 200 nucleotides can be transcribed (Chao et al., 1999).

Unknowing whether the pGL4.73 template plasmid's sequence was sufficiently GU/U-rich for reliable cleavage-and-polyadenylation to occur, and whether transcribed native downstream sequences had other expression-relevant effects prior to cleavage, three 3′UTR types were reproduced with extensions of between 500bp and 1kb from downstream of the cleavage-and-polyadenylation site gene template. Plasmid constructs created after this point in time were similarly extended, with the aim of preserving the native cleavage-and-polyadenylation context.

For all forward-orientation 3′UTR types compared to extended versions, the data point spread within the unextended 3′UTR type exceeded the distance between the unextended and extended 3′UTR types.

For the backward-orientation flex-cassette alone, the absence of the downstream 3′UTR produced a drop in expression of 4-fold or greater, suggesting that the majority of expression originates from that downstream 3′UTR, as intended. That said, this expression level (2^{-6} of (+SV40)) was still 8-fold greater than the lowest expression observed (+mGDF11, discussed in subsection 6.3, with expression 2^{-9} of (+SV40)), which bears further investigation such as through RNA sequencing of generated transcripts.

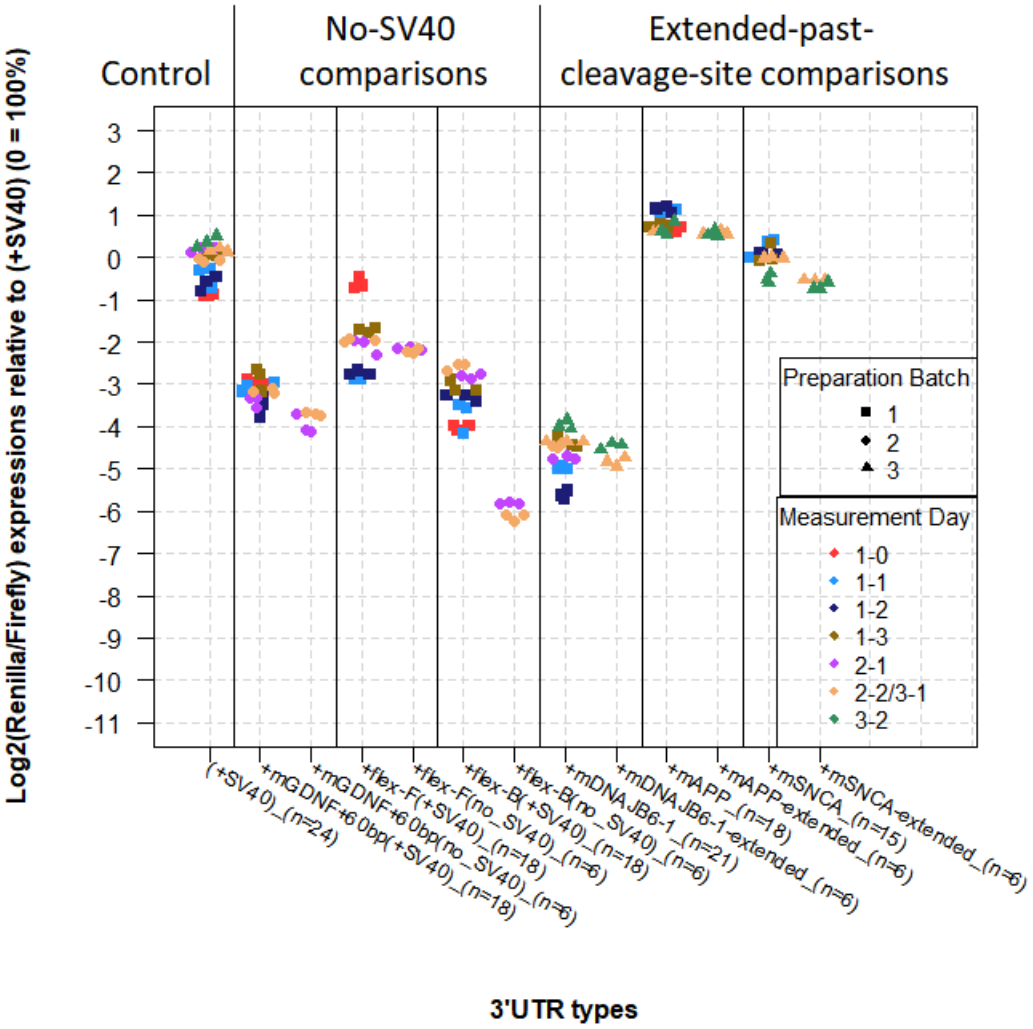

Figure 12. Readthrough-influence overall test. These are the effects on protein expression of a 3′UTR's downstream sequence. The backwards-orientation flex-B sequence, intended to be read through without signalling transcript cleavage, suffers an approximately 8-fold drop when the (+SV40) 3′UTR following it is removed. The mGDNF and flex-F sequence, which having forwards-orientation are intended to signal transcript cleavage, have a small change in expression if any after the same excision. The addition of genomic DNA following the transcript cleavage site for mDNAJB6-1, mAPP, and mSNCA results in a small change in expression if any. Together this suggests a low influence of downstream sequences on forward-direction 3′UTRs. All measurement day and 3′UTR type abbreviations are listed and described in Appendix A.

6.4.3. Lox-influence test

Having established that 3′UTR conditional replacement largely works as intended, the other aspect to investigate was whether the presence of the lox sites was responsible for the lower protein expression from +flex-B(+SV40), perhaps through hairpin formation in the plasmid. To do this, bGHpA in backward- and forward-orientation was tested with a loxP and lox5171 site on each side (‘flex’), with the outermost lox site removed from each side (‘1lox’), and with all lox sites removed (‘0lox’).

Kakoki et al.'s work (Kakoki et al., 2004) does suggest that a loxP site can lower expression, though as their work was done through genomic alteration and the purpose of this test was to investigate a discrepancy between genomic effects and plasmid effects, this is of limited relevance unless a ‘+flex-B+mGDNF’ 3′UTR type did not display a lower expression than a ‘+mGDNF’ 3′UTR type.

As shown in Figure 13, removal of all lox sites flanking the backwards-orientation bGHpA sequence fully restored (+SV40) expression, indicating that the lox sites—an intrinsic part of using a flex-cassette for conditional knockup—were responsible for the lowered expression levels observed. The expression from the forward-orientation bGHpA was also improved by the removal of lox sites, though to a lesser degree—perhaps due to the lox sites being closer to the end of the produced mRNA transcript than in backward-orientation cases for which the downstream SV40 3′UTR is not cleaved away. As the GDNF 3′UTR is over 2 kb long (see Appendix A), this hypothesis bears further investigation: for instance, through the comparison of lox sites inserted at different positions in a long 3′UTR sequence.

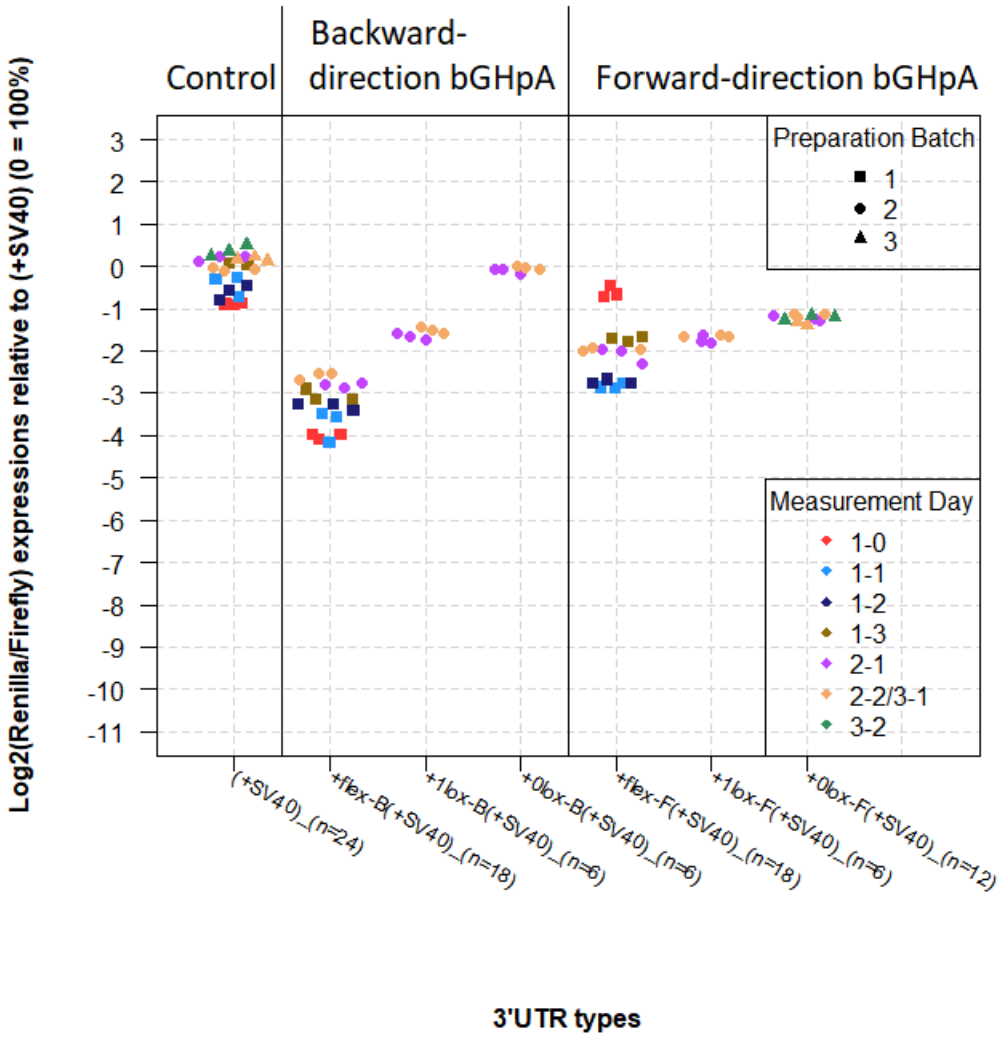

Figure 13. Lox-influence conditional knockup test. These data points are bGHpA in the backward- and forward-orientation with the flex-cassette flanking loxP and lox5171 sites progressively stripped away. (+SV40) is side-by-side with +flex-B(+SV40), showing that the backward-orientation flex-cassette addition decreases expression approximately 8-fold; stepwise removal of all lox sites (+1lox-B(+SV40), +0lox-B(+SV40)) then shows how expression can be completely restored while the backward-orientation bGHpA sequence remains. Expression from forward-orientation bGHpA (+flex-F(+SV40), +1lox-F(+SV40), +0lox-F(+SV40)) also shows improvement. All measurement day and 3′UTR type abbreviations are listed and described in Appendix A.

6.5. Comparison of different co-transfection ratios

Promega's dual-luciferase assay manual warns that co-transfected plasmids with identical promoters can suffer from trans effects, and recommends a control:reporter co-transfection molar ratio between 1:10 and 1:50 (Dual-Luciferase Reporter Assay System manual subsection 3.B.). pGL4.13 and pGL4.73 both use the SV40 enhancer and promoter, and while measurement days 1-0, 1-1, and 1-2 were conducted with a 1:11.8 control:reporter ratio, all other measurement days were conducted with a 1:10 ratio, right at the edge of the recommended range.

For these 1:11.8 measurement days, +mAPP seemed to display a high expression relative to (+SV40) (Fig. 14), yet a near-equal expression to (+SV40) in later measurements, including for +mAPP-extended. To check this, (+SV40) and +mAPP were first tested in adjacently-measured triplets with control:reporter co-transfection molar ratios of 1:10 and 1:25 as part of measurement day 3-2 (Fig. 14).

As the +mAPP expression appeared a little higher than the (+SV40) for the 1:25 co-transfections, a more thorough comparison was carried out on measurement day 4-0 with ratios of 1:10, 1:20, 1:40, 1:80, and 1:60. Measurement ratios were scaled according to each co-transfection ratio and compared in a single chart (Fig. 15). This showed no divergent-expression trend for different co-transfection ratios, and so it was concluded that initially high +mAPP relative expression was due to differing plasmid degradation or other initial unreliability, rather than due to the co-transfection ratio.

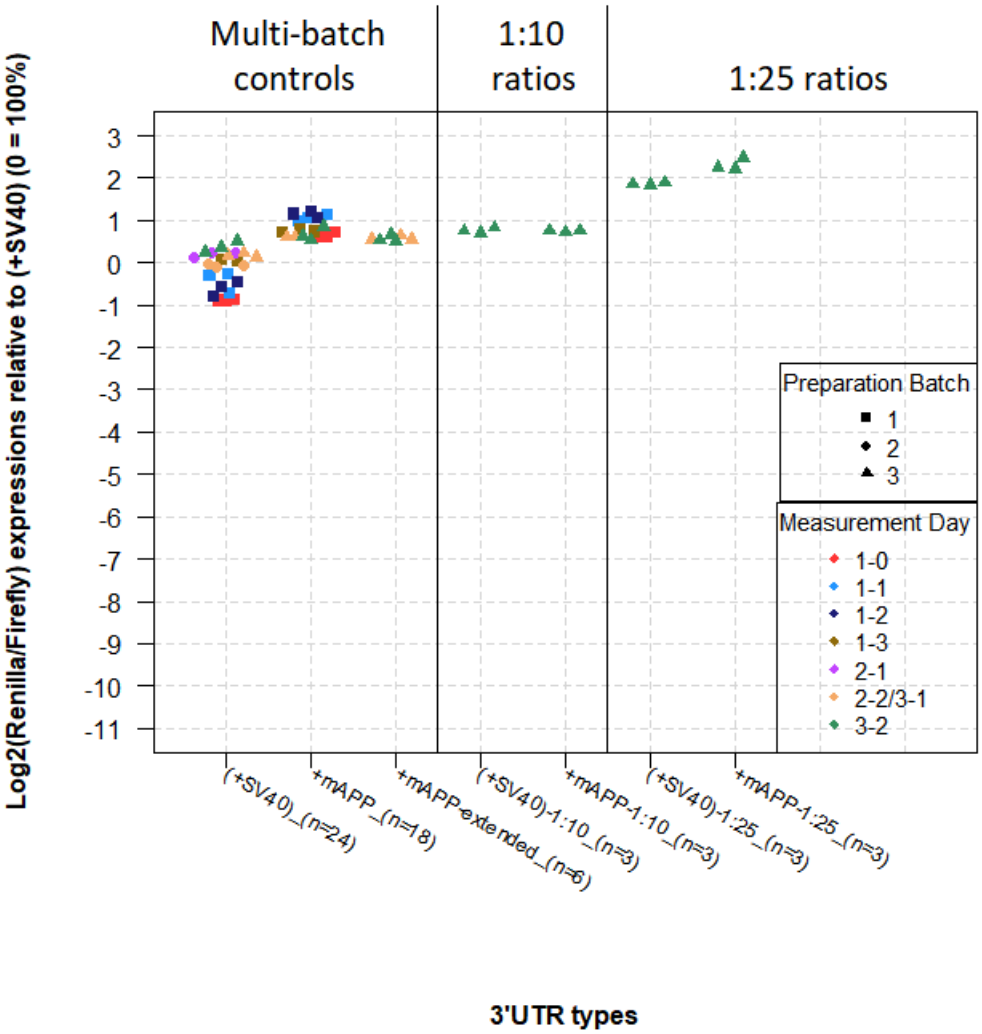

Figure 14. A preliminary test of the effect of different co-transfection ratios. For a 1:10 control:reporter plasmid ratio, (+SV40) and +mAPP presented similar expression levels, whereas for a 1:25 ratio the expression level of +mAPP appeared a little higher than that of (+SV40). All measurement day and 3′UTR type abbreviations are listed and described in Appendix A.

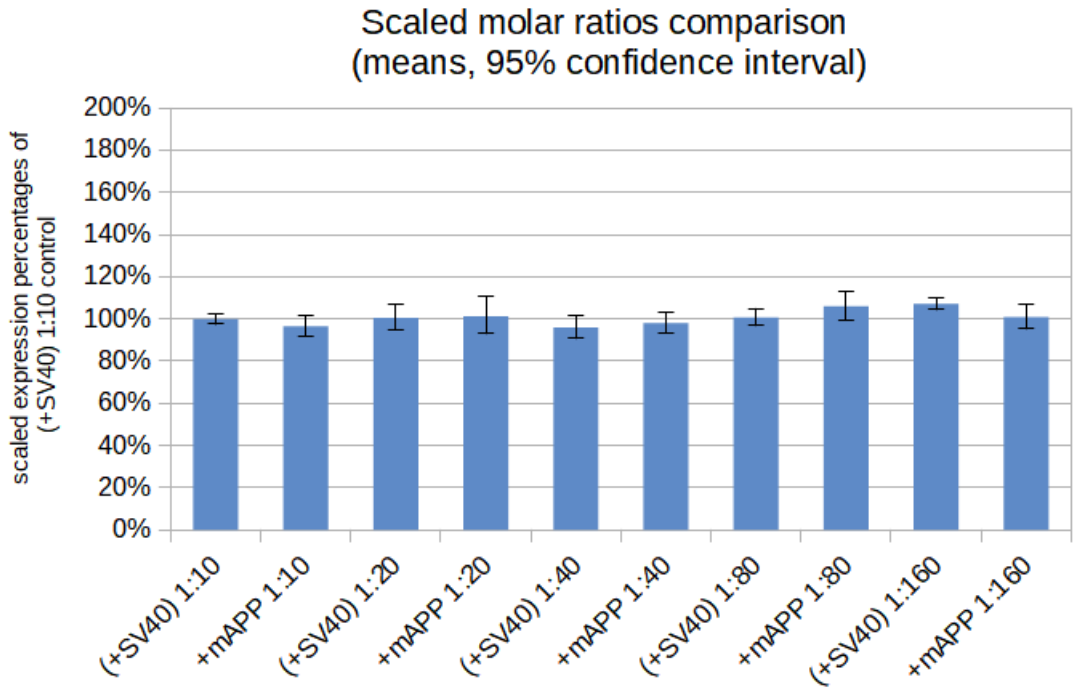

Figure 15. A further test of the effect of different co-transfection ratios. In a small batch solely for ratio-testing, plasmids (+SV40) and +mAPP were compared to each other at control:reporter co-transfection ratios of 1:10, 1:20, 1:40, 1:80, and 1:160, and the displayed results were scaled according to the ratio used. All displayed results were similar, with no apparent expression difference emerging between (+SV40) and +mAPP at more extreme co-transfection ratios. Means and 95% confidence intervals were calculated from log2(Renilla/Firefly) ratios as 1.96*{Sample standard deviation}/sqrt({Sample size}), with {Sample size}=3 for all ten samples, and charted in LibreOffice Calc. All measurement day and 3′UTR type abbreviations are listed and described in Appendix A.

6.6. Candidates for bGHpA flex-cassette conditional knockup

Lox-site-present in vitro behaviour was seemingly inconsistent with in vivo behaviour, but as the +1lox-F expression relative to +mGDNF+60bp(+SV40) was seemingly consistent with in vivo mouse kidney GDNF conditional knockup results (subsection 6.4.1), in vivo mouse models could in future be constructed to examine whether in vivo forward-orientation predictions are reliable.

For the genomic 3′UTRs used in this project, conditional knockup candidates are those for which the expression level was measured to be lower than that of +1lox-F, namely the 3′UTRs of +mDNAJB6-3, +mPARK2-1, +mMME, and +mDNAJB6-1 (Fig. 16). Of these +mPARK2-1 belonged to the neuron survival gene group and the others to the anti-aggregation gene group, and +mDNAJB6-1 displayed the lowest expression relative to +1lox-F.

For mouse DnaJ heat shock protein family (Hsp40) member B6 (DNAJB6) in particular, transcript-specific qPCR primers should be used when measuring in vivo mRNA levels to avoid confusion with expression from other transcripts' 3′UTRs.

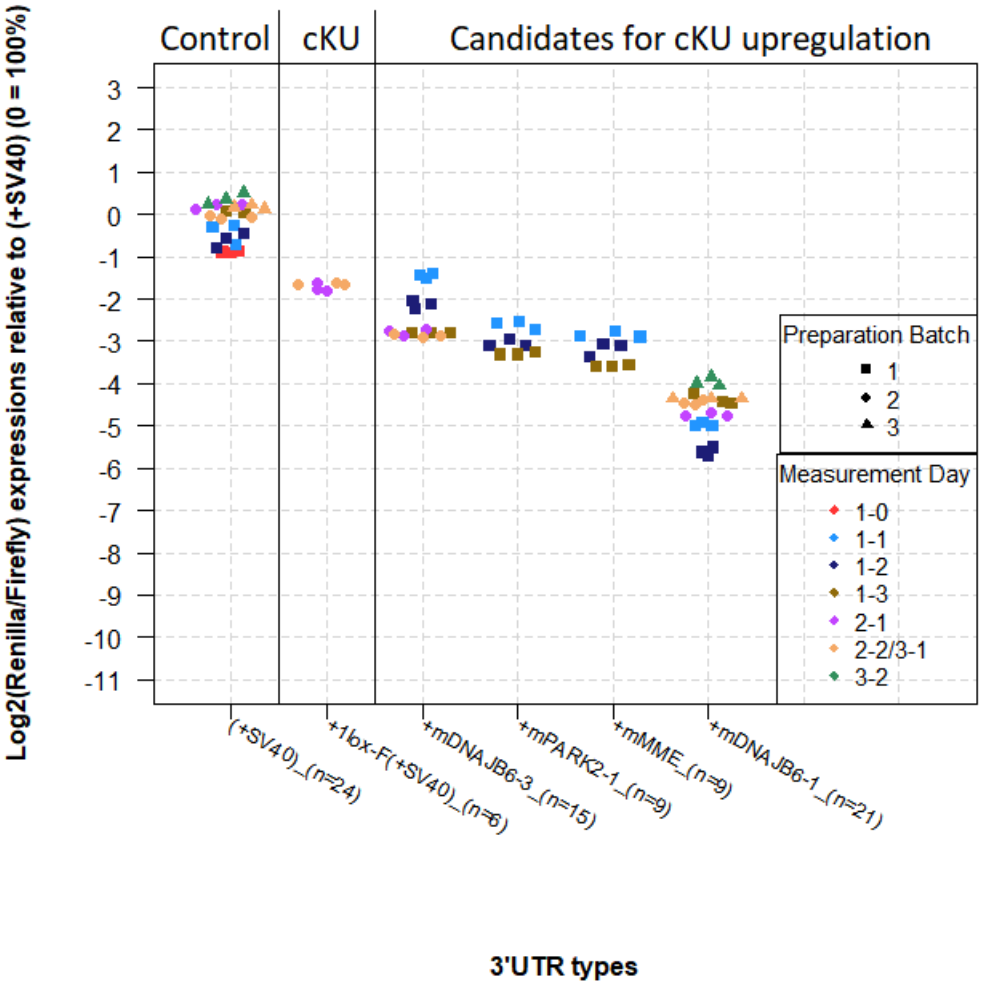

Figure 16. Candidates for flex-cassette conditional knockup. These are the measured 3′UTR types with potentially-lower expression than +1lox-F, the 3′UTR type representing an excision-flipped bGHpA flex-cassette inserted upstream of a native 3′UTR. The four candidate 3′UTR types are arranged in approximate order of decreasing expression, and though +mDNAJB6-3 includes measured relative expressions close to and above those of +1lox-F, these are from the 1-0/1-1/1-2 measurement days viewed as less reliable than the later measurement days (subsections 6.1 and 6.3). All measurement day and 3′UTR type abbreviations are listed and described in Appendix A.

6.7. Alternatives to bGHpA for flex-cassette 3′UTR replacement

While 3′UTR types with lower expression than +1lox-F represent candidates for conditional knockup with the bGHpA flex-cassette, 3′UTR types with higher expression than +0lox-F instead represent candidates to use in a flex-cassette as an alternative 3′UTR to bGHpA.

A significant factor however when considering bGHpA replacement is that of 3′UTR length, in that viral vectors and other DNA delivery methods typically have limited sequence capacity. In practice, the longer the replacement 3′UTR, the less capacity is available for protein coding sequences or gene control elements. While mouse model creation is not limited by this, some thought could still be spent on vector requirements for eventual adult treatment.

Though the bGHpA sequence has a length of only 297 bp, candidate mouse 3′UTRs in this project were those of +mSNCA (556 bp), +mDNAJB6-2 (671 bp), +mAPP (895 bp), +mACTB-extended (683+530 bp), and +mDNAJC6 (2226 bp), all more than twice the length of bGHpA (Fig. 17; Appendix A).

The (+SV40) (256 bp) and +WPRE+bGH (1170 bp) sequences also showed comparatively high expressions in this assay, but viral-origin sequences are at risk of undergoing eventual epigenetic silencing in mammalian cells, which conflicts with the desired purpose of long-term medical treatment, and WPRE which showed the highest expression in the form of +WPRE+bGH is unusable in some cell types (Klein et al., 2006).

(+SV40) though short in length is also unusable as a flex-cassette candidate due to being the shared 3′UTR region of two opposite-direction SV40 genes, SV40gp5 and SV40gs1. The pGL4.73 (+SV40) 256 bp 3′UTR consists of 24 bp SV40-unrelated sequence, 56 bp forward-orientation SV40gp5 coding sequence and stop codon, 97 bp shared-3′UTR-region in the SV40gp5 direction, 69 bp backward-orientation SV40gs1 coding sequence and stop codon, and 10 bp SV40-unrelated sequence (according to NCBI's reference sequence NC_001669.1 (Fig. 18)). While a flex-cassette replacement 3′UTR should cause cleavage-and-polyadenylation in the forward-orientation and not in the backward-orientation, (+SV40) thus does so for both orientations.

Notably, the 145 bp SV40 3′UTR sequence successfully used by Kakoki et al. (Kakoki et al., 2004) is in the opposite orientation and shares 130 bp of overlap with the pGL4.73 (+SV40) 3′UTR, consisting of 153 bp forward-orientation SV40gs1 coding sequence and stop codon, together with 61 bp shared-3′UTR-region in the SV40gs1 direction. While not containing the full shared-3′UTR-region, the Kakoki SV40 3′UTR still contains SV40gp5's polyadenylation signal and one of SV40gs1's two; all three polyadenylation signals are closer to the stop codon of the opposite direction from them than to their same-direction stop codon, and so preventing cleavage-and-polyadenylation through sequence truncation appears infeasible.

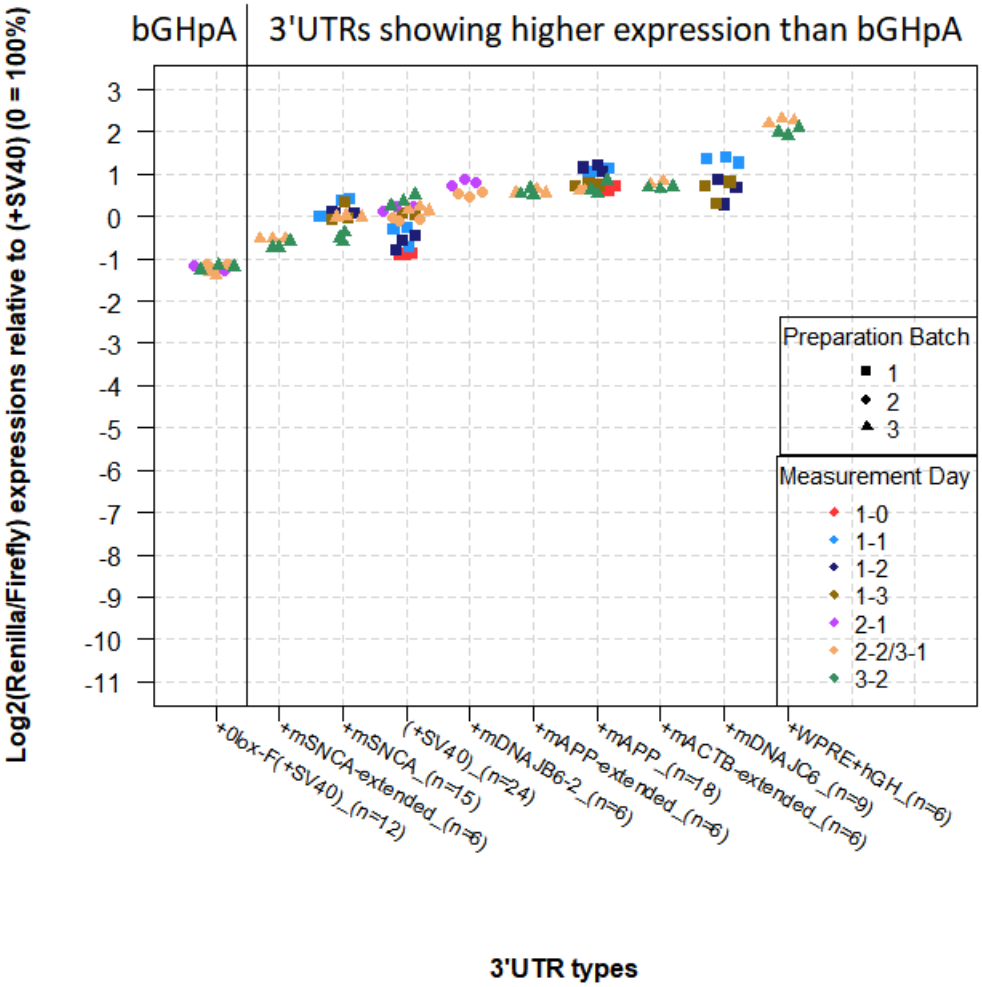

Figure 17. Flex-cassette alternatives to bGHpA for 3′UTR replacement. These are the 3′UTR types with potentially-higher expression than +0lox-F, the 3′UTR type which is forward-orientation bGHpA. The candidate 3′UTR types are arranged in approximate order of increasing expression. All measurement day and 3′UTR type abbreviations are listed and described in Appendix A.

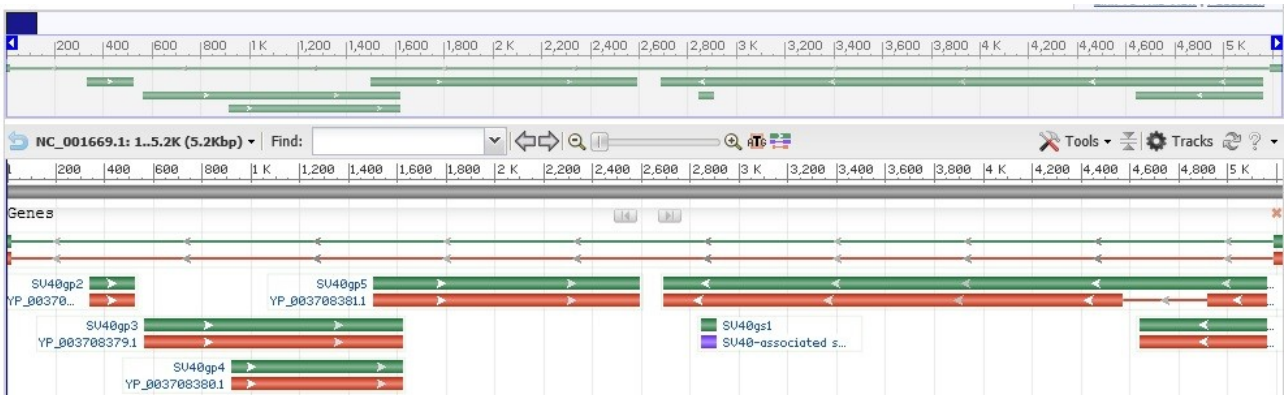

Figure 18. NCBI SV40 3′UTR context, a screenshot from https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=23950 . This is a section of the complete SV40 genome in NCBI, accession number NC_001669.1, that shows the opposite-directions SV40gp5 and SV40gs1 shared 3′UTR region (the 222 bp region from base 2,538 through base 2,759) used in plasmid pGL4.73.

6.8. Possible workflow improvements for future experiments

If further in vitro experiments were carried out prior to or in parallel with in vivo comparisons, there are a number of changes, additions or deeper investigations that could be recommended.

6.8.1. Mouse cell line

In the current workflow, mouse 3′UTRs were tested in the human embryonic kidney cell line HEK-293, for the reasons that HEK-293 was on-hand, grew quickly, and had good transfection efficiency. Mouse 3′UTRs were used with in vivo mouse models in mind, but for more assured emulation of the native environment's RNA regulation a mouse cell line should be used instead. Alternatively, the microRNA expression profile of a non-mouse cell line used could be compared directly to the microRNA expression profile of the intended-treatment-target mouse cell type, particularly paying attention to microRNAs with binding sites present in the affected sequences, perhaps as indicated by TargetScan Mouse (http://www.targetscan.org/mmu_71/).

Further, while immortalised cell lines are always—through aneuploidy, mutation, or otherwise—very different in behaviour from their original cell types, where upregulation is being considered in the context of treatment for neurodegeneration in brain cells, use of a brain cell line could closer approach the environment of interest. This is particularly in light of the different degrees of upregulation derived from bGHpA flex-cassette knockup of GDNF in mouse kidney versus striatum (Fig. 3), since microRNA expression varies according to cell type (Shivdasani, 2006). For example, the Neuro-2a cell line originates from mouse neuroblast cells. Alternatively, several different mouse or mouse and human cell lines of differing origin could be used to estimate the influence of cell type on protein expression for each 3′UTR.

Further still, in the current workflow some 3′UTRs were tested which have no homologous 3′UTR sequence in the homologous human gene. As an example, for the mouse DNAJB6 gene only the 3′UTR for isoform c (‘DNAJB6-2’, reference sequence NM_011847.4) is homologous to a human DNAJB6 3′UTR, that of isoform b (reference sequence NM_005494.2): 574/698 identities, according to NCBI's Align Sequences Nucleotide BLAST blastn algorithm (https://blast.ncbi.nlm.nih.gov/Blast.cgi). If testing for the sake of future human medical treatment rather than only as proof-of-concept, mouse genomic 3′UTR types tested could be limited to those with human homologues. While the mouse 3′UTR types were tested in mouse cell line contexts, their human homologues could be tested in human cell line contexts, ideally both cell lines being of the same intended-treatment-target cell type.

6.8.2. Plasmid condition and concentration measurement

Though plasmid degradation was observed (subsection 6.1) and guessed to be caused by vortexing, this was not tested further. Vortexing of different aliquots of the same plasmid at different speeds and for different lengths of time could be carried out to determine whether this can in fact produce significant plasmid nicking and fragmentation, and if so what limits on vortex-mixing would acceptably mitigate this.

Another matter potentially able to affect plasmid condition is that of the elution buffer used. As noted in Materials and Methods, extracted plasmids were stored in and diluted with sterile water rather than the plasmid extraction kit's elution buffer. This was to ensure a plentiful supply for dilutions, as well as motivated by observing no immediate problems with NanoDrop-measured elution efficiency or effective transfection efficiency. However, irreversible structural changes to DNA diluted in the absence of NaCl have been reported (Nakayama et al., 2016), and so this is another hypothetical contributor to the observed degradation. This could likewise be tested by comparison of whether elution and dilution in elution buffer instead of sterile water can mitigate any repeatable degradation observed.

Finally, in the same gel in which degradation was observed (Fig. 7), the unreliability of NanoDrop-measured elution concentrations is displayed, which could skew intended co-transfection ratios. If inaccurate concentration measurements are due to chaotropic salt interference, then more accurate measurements might be obtainable from a Qubit fluorometer, which measures DNA-bound dye instead of 260 nanometer UV absorbance.

It could also be tested what concentrations a Qubit fluorometer reports for degraded DNA. If degradation consistently reduced DNA to unmeasured free nucleotides, rather than to normally-measured small fragments, then Qubit concentration measurement could be used immediately before co-transfection. This would allow preparation of correct-ratio co-transfection mixtures even from partially-degraded plasmid preparations, rather than requiring the preparation of mixes for all of a batch's measurement days immediately after plasmid extraction.

Alternatively, several mixes with different stochastic pipetting errors could be simultaneously prepared for each 3′UTR immediately after plasmid extraction, to reduce pipetting error influence on a 3′UTR type measured in only one batch.

6.8.3. RNA transcript sequence verification

Noted in subsection 6.4.2, the +flex-B(no_SV40) 3′UTR type showed an expression level much higher than that of +mGDF11, even though +flex-B(no_SV40) had been thought to be the least viable of all 3′UTR types tested. RNA sequencing would indicate what RNA transcripts are being produced from these two 3′UTR types, offering clues as to why one's relative expression is higher than the other (such as undesired cleavage-and-polyadenylation within the backward-orientation flex-cassette, which might then be preventable).

More generally, this could reveal for all measured 3′UTR types any undesired downstream plasmid sequence present in a polyadenylated mRNA transcript, such as when a reference sequence has included an insufficient length of a genomic 3′UTR.

In practice, Northern blotting with a probe specific to the Renilla luciferase coding sequence could be a less costly approach to check all tested 3′UTR types for surprising transcript lengths, prior to more in-depth investigation.

6.8.4. Flex-cassette internal sequence

As described in subsection 6.7, other sequences might offer more reliable results in a gene knockup flex-cassette than bGHpA. A choice of different complete 3′UTRs would further allow fine adjustment when a specific new expression level is desired.

Other than influencing protein expression levels through the recruitment of proteins which affect mRNA stability, sequences in the 3′UTR can also exert different forms of posttranscriptional control by recruiting proteins with other roles, such as intra-cell localisation and mRNA-dependent protein-protein interactions (Tushev et al., 2018; Mayr, 2019). There are even reports of 3′UTRs responsible for transcriptional silencing or enhancing of their own genes (Gomez-Benito et al., 2011; Jash et al., 2012). In order to modify only protein expression levels and not protein behaviour, 3′UTR modifications would need to only affect sequences which influence mRNA stability.

A way to achieve the same flexibility while meeting this requirement could instead be the use of multiple mRNA-half-life-increasing sequences, such as a cytoplasmic polyadenylation element or WPRE sequence, to impart a scalable incremental expression increase. This would preserve microRNA cell-type-specific regulation, as well as any sequences which affect protein behaviour outside mRNA stability.

6.8.5. Statistical analysis method

In this project, data-points were presented directly in strip-chart scatterplots in lieu of conducting a numerical analysis. It can be speculated that there should be a way to meaningfully analyse the recorded measurements, and if so then this would be a preferable method for future studies.

It would not be appropriate to redesign the experimental procedures in order to improve estimation simplicity by narrowing data collection to a single set of conditions (such as carrying out measurements for only one large co-transfected sample on a single measurement day). This would not remove the bias from stochastic errors in experiment preparation, and instead would remove points of reference for identifying and partly correcting for such bias. Further, carrying out small lots of measurements on different days scales better for larger amounts of data: each small lot of measurements benefits from assay substrates in approximately the same condition, as well as the full concentration of the experimenter. For a single large lot of measurements, the first samples would have fresh assay substrates and a well-rested experimenter, but the last samples would instead have less-fresh substrates and a fatigued experimenter more prone to careless errors.

Though logarithmisation of measurement ratios appears appropriate for photomultiplier tube values, this relies on an assumption of normally-distributed measurement values after logarithmisation, which bears further consideration. Normality testing of logarithmised measurement ratio sample residuals (subsection 6.2.3) suggested that the residuals were not normally-distributed, but the implications of this are not clear unless first determining whether the residuals of normal distributions with different means are expected to be normally-distributed when outside the context of a linear regression analysis. The inability to assume that samples possessed a common population variance (subsection 6.2.2) is similarly concerning.

If recorded measurements are of different distributions from those expected, their actual properties must be learned for those expectations to be adjusted and the data appropriately handled. If recorded measurements are instead indeed of the expected distributions, then analysis methods adequate to conclude this are desirable.

7. Conclusions

The largest contribution of this project has been a refining of the workflow used, namely the identification of problematic areas and the development of ways to address them, as well as the determination of future areas that could be focused on to answer remaining questions. That said, the overall conclusion for the specific results obtained is that they are insufficiently conclusive and so require further studies.

In comparing different genomic 3′UTRs to each other in a known environment, no overarching trends in the gene categories focused on were evident.

If the in vivo human embryonic kidney cell line results had been wholly consistent with the in vivo mouse kidney GDNF conditional knockup results, then it would have been a point in favour of the reliability of this in vitro method for predicting the in vivo effects of 3′UTR modification. Even so, further in vivo comparisons would have been appropriate in case that consistency had been a rarely-occurring coincidence. As there was only partial consistency (subsection 6.4.1), the impression obtained is instead one of mixed hope and doubt.

For the impression which gave hope, that hope can be put to the test through more in vivo mouse models for comparison, using several genes and tissues, while also expanding the in vitro tests to other cell lines.

For the impression which gave doubt, that doubt can be put to the test through either a bottom-up approach, a top-down approach, or ideally both. The bottom-up approach is to test the effect of lox sites on protein expression according to their distance from the 3′UTR's end. The top-down approach is to further pursue the creation of a ‘+flex-B+mGDNF’ 3′UTR type despite the long sequence length, perhaps through long multiple fusion of PCR products (Shevchuk et al., 2004), and see whether its expression matches that of a ‘+mGDNF’ 3′UTR type. Otherwise, if different in vivo results consistently matched in vitro results for identical 3′UTR sequences, the mouse GDNF instance could be temporarily treated as a rare inconsistency until further investigation were possible.

Other than these approaches for testing whether this in vitro method is usably predictive of in vivo results, there remain a number of avenues of possible workflow improvement (subsection 6.8). Even if it were instead not predictive of in vivo results, this work could perhaps still be of use in improvement of in vitro industrial protein production.

8. References

n.d. and n.n. refer to ‘no day’ and ‘no name’ respectively, for when the date of writing or author's name are unknown.

Barde MP, Barde PJ. (2012) What to use to express the variability of data: Standard deviation or standard error of mean?. *Perspect Clin Res*. 3(3):113-116. doi: 10.4103/2229-3485.100662

Ben-Haim MS, Kanfi Y, Mitchell SJ, Maoz N, Vaughan KL, Amariglio N, Lerrer B, de Cabo R, Rechavi G, Cohen HY. (2018) Breaking the Ceiling of Human Maximal Life span. *J Gerontol A Biol Sci Med Sci.* 73(11):1465-1471. doi: 10.1093/gerona/glx219. PubMed PMID: 29121176; PubMed Central PMCID: PMC6454488.

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT. (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem.* 55(4):611-22. doi: 10.1373/clinchem.2008.112797. PubMed PMID: 19246619.

Chao LC, Jamil A, Kim SJ, Huang L, Martinson HG. (1999) Assembly of the cleavage and polyadenylation apparatus requires about 10 seconds in vivo and is faster for strong than for weak poly(A) sites. *Mol Cell Biol.* 19(8):5588-600. PubMed PMID: 10409748; PubMed Central PMCID: PMC84411.

Chen L, Morrow JK, Tran HT, Phatak SS, Du-Cuny L, Zhang S. (2012) From laptop to benchtop to bedside: structure-based drug design on protein targets. *Curr Pharm Des.* 18(9):1217-39. PubMed PMID: 22316152; PubMed Central PMCID: PMC3820560.

Chiti F, Dobson CM. (2017) Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu Rev Biochem.* 86:27-68. doi: 10.1146/annurev-biochem-061516-045115. PubMed PMID: 28498720.

Cieri F, Esposito R. (2019) Psychoanalysis and Neuroscience: The Bridge Between Mind and Brain. *Front Psychol.* 10:1790. doi: 10.3389/fpsyg.2019.01983. PubMed PMID: 31555159; PubMed Central PMCID: PMC6724748.

Cohen, J. (1994) The earth is round (p < .05). *American Psychologist.* 49(12), 997–1003. doi: 10.1037/0003-066X.49.12.997.

Corrigan F, Vink R, Blumbergs PC, Masters CL, Cappai R, van den Heuvel C. (2012) Characterisation of the effect of knockout of the amyloid precursor protein on outcome following mild traumatic brain injury. *Brain Res.* 1451:87-99. doi: 10.1016/j.brainres.2012.02.045. PubMed PMID: 22424792.

Cosacak MI, Bhattarai P, Reinhardt S, Petzold A, Dahl A, Zhang Y, Kizil C. (2019) Single-Cell Transcriptomics Analyses of Neural Stem Cell Heterogeneity and Contextual Plasticity in a Zebrafish Brain Model of Amyloid Toxicity. *Cell Rep.* 27(4):1307-1318.e3. doi: 10.1016/j.celrep.2019.03.090. PubMed PMID: 31018142.

Ellenbroek B, Youn J. (2016) Rodent models in neuroscience research: is it a rat race? *Dis Model Mech.* 9(10):1079-1087. PubMed PMID: 27736744; PubMed Central PMCID: PMC5087838.

Emerson M. (2010) Refinement, reduction and replacement approaches to in vivo cardiovascular research. *Br J Pharmacol.* 161(4):749-754. doi: 10.1111/j.1476-5381.2010.00959.x

Gao X, Carroni M, Nussbaum-Krammer C, Mogk A, Nillegoda NB, Szlachcic A, Guilbride DL, Saibil HR, Mayer MP, Bukau B. (2015) Human Hsp70 Disaggregase Reverses Parkinson's-Linked α-Synuclein Amyloid Fibrils. *Mol Cell.* 59(5):781-93. doi: 10.1016/j.molcel.2015.07.012. PubMed PMID: 26300264; PubMed Central PMCID: PMC5072489.

Garringer HJ, Sammeta N, Oblak A, Ghetti B, Vidal R. (2016) Amyloid and intracellular accumulation of BRI(2). *Neurobiol Aging.* 52:90-97. doi: 10.1016/j.neurobiolaging.2016.12.018. PubMed PMID: 28131015; PubMed Central PMCID: PMC5359036.

Geary, R. (1936). The Distribution of "Student's" Ratio for Non-Normal Samples. *Supplement to the Journal of the Royal Statistical Society.* 3(2), 178-184. doi: 10.2307/2983669.

Greenland S, Senn SJ, Rothman KJ, et al. (2016) Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol*. 31(4):337-350. doi: 10.1007/s10654-016-0149-3

Gu S, Jin L, Zhang F, Sarnow P, Kay MA. (2009) Biological basis for restriction of microRNA targets to the 3' untranslated region in mammalian mRNAs. *Nat Struct Mol Biol.* 16(2):144-50. doi: 10.1038/nsmb.1552. PubMed PMID: 19182800; PubMed Central PMCID: PMC2713750.

Gurevich EV, Gurevich VV. (2015) Beyond traditional pharmacology: new tools and approaches. *Br J Pharmacol.* 172(13):3229-41. doi: 10.1111/bph.13066. PubMed PMID: 25572005; PubMed Central PMCID: PMC4500362.

Gomez-Benito M, Loayza-Puch F, Oude Vrielink JAF, Odero MD, Agami R. (2011) 3'UTR-mediated gene silencing of the Mixed Lineage Leukemia (MLL) gene. *PLoS One.* 6(10):e25449. doi: 10.1371/journal.pone.0025449. PubMed PMID: 21998658; PubMed Central PMCID: PMC3187771. Erratum doi: 10.1371/annotation/d39cb963-9277-462f-8070-6409d6fcf484.

Gründemann D, Schömig E. (1996) Protection of DNA during preparative agarose gel electrophoresis against damage induced by ultraviolet light. *Biotechniques.* 21(5):898-903. PubMed PMID: 8922632.

Hahn MW, Wray GA. (2002) The g-value paradox. *Evol Dev.* 4(2):73-5. PubMed PMID: 12004964.

Ibáñez CF, Andressoo JO. (2016) Biology of GDNF and its receptors - Relevance for disorders of the central nervous system. *Neurobiol Dis.* 97(Pt B):80-89. doi: 10.1016/j.nbd.2016.01.021. PubMed PMID: 26829643.

Jalkanen AL, Coleman SJ, Wilusz J. (2014) Determinants and implications of mRNA poly(A) tail size--does this protein make my tail look big? *Semin Cell Dev Biol.* 34:24-32. doi: 10.1016/j.semcdb.2014.05.018. PubMed PMID: 24910447; PubMed Central PMCID: PMC4163081.

Jash A, Yun K, Sahoo A, So JS, Im SH. (2012) Looping mediated interaction between the promoter and 3' UTR regulates type II collagen expression in chondrocytes. *PLoS One.* 7(7):e40828. doi: 10.1371/journal.pone.0040828. PubMed PMID: 22815835; PubMed Central PMCID: PMC3397959.

Kakoki M, Tsai YS, Kim HS, Hatada S, Ciavatta DJ, Takahashi N, Arnold LW, Maeda N, Smithies O. (2004) Altering the expression in mice of genes by modifying their 3' regions. *Dev Cell.* 6(4):597-606. PubMed PMID: 15068798.

Kass RE, Raftery AE. (1995) Bayes Factors. *Journal of the American Statistical Association.* 90:430, 773-795. doi: 10.1080/01621459.1995.10476572.

Knight K. (2000) *Mathematical Statistics.* New York, United States: Chapman and Hall.

Kissick DJ, Muir RD, Simpson GJ. (2010) Statistical treatment of photon/electron counting: extending the linear dynamic range from the dark count rate to saturation. *Anal Chem.* 82(24):10129-34. doi: 10.1021/ac102219c. PubMed PMID: 21114249; PubMed Central PMCID: PMC3025609.

Klein R, Ruttkowski B, Knapp E, Salmons B, Günzburg WH, Hohenadl C. (2006) WPRE-mediated enhancement of gene expression is promoter and cell line specific. *Gene.* 372:153-61. PubMed PMID: 16488559.

Kumar A, Kopra J, Varendi K, Porokuokka LL, Panhelainen A, Kuure S, Marshall P, Karalija N, Härma MA, Vilenius C, Lilleväli K, Tekko T, Mijatovic J, Pulkkinen N, Jakobson M, Jakobson M, Ola R, Palm E, Lindahl M, Strömberg I, Võikar V, Piepponen TP, Saarma M, Andressoo JO. (2015) GDNF Overexpression from the Native Locus Reveals its Role in the Nigrostriatal Dopaminergic System Function. *PLoS Genet.* 11(12):e1005710. doi: 10.1371/journal.pgen.1005710. Erratum (2016) in: *PLoS Genet.* 12(1):e1005808. PubMed PMID: 26681446; PubMed Central PMCID: PMC4682981.

Lee PY, Costumbrado J, Hsu CY, Kim YH. (2012) Agarose gel electrophoresis for the separation of DNA fragments. *J Vis Exp.* (62). doi: 10.3791/3923. PubMed PMID: 22546956; PubMed Central PMCID: PMC4846332.

Maskri L, Zhu X, Fritzen S, et al. (2004) Influence of different promoters on the expression pattern of mutated human alpha-synuclein in transgenic mice. *Neurodegener Dis*. 1(6):255-265. doi: 10.1159/000085064

Mätlik K, Olfat S, Garton D, Montaño A, Turconi G, Porokuokka L, Panhelainen A, Schweizer N, Kopra J, Cowlishaw MC, Piepponen PT, Zhang FP, Sipilä P, Jakobsson J, Andressoo JO. (2019). Gene Knock Up via 3′UTR editing to study gene function in vivo. doi: 10.1101/775031.

Mayr C. (2019) What Are 3' UTRs Doing? *Cold Spring Harb Perspect Biol.* 11(10). pii: a034728. doi: 10.1101/cshperspect.a034728. PubMed PMID: 30181377.

McCray BA, Taylor JP. (2008) The role of autophagy in age-related neurodegeneration. *Neurosignals.* 16(1):75-84. PubMed PMID: 18097162.

Mignone F, Gissi C, Liuni S, Pesole G. (2002) Untranslated regions of mRNAs. *Genome Biol.* 3(3):REVIEWS0004. PubMed PMID: 11897027; PubMed Central PMCID: PMC139023.

Milholland B, Suh Y, Vijg J. (2017) Mutation and catastrophe in the aging genome. *Exp Gerontol.* 94:34-40. doi: 10.1016/j.exger.2017.02.073. PubMed PMID: 28263867; PubMed Central PMCID: PMC5480213.

Murphy KP. (2012) *Machine Learning: A Probabilistic Perspective.* Cambridge, United States: MIT Press.

Nakayama Y, Yamaguchi H, Einaga N, Esumi M. (2016) Pitfalls of DNA Quantification Using DNA-Binding Fluorescent Dyes and Suggested Solutions. *PLoS One.* 11(3):e0150528. doi: 10.1371/journal.pone.0150528. PubMed PMID: 26937682; PubMed Central PMCID: PMC4777359.

NEB (n.d.) Cleavage Close to the End of DNA Fragments *Retrieved March 30th, 2020 from* https://international.neb.com/tools-and-resources/usage-guidelines/cleavage-close-to-the-end-of-dna-fragments

Nguyen D, Xu T. (2008) The expanding role of mouse genetics for understanding human biology and disease. *Dis Model Mech.* 1(1):56-66. doi: 10.1242/dmm.000232. PubMed PMID: 19048054; PubMed Central PMCID: PMC2561976.

Oppenheim A. (1981) Separation of closed circular DNA from linear DNA by electrophoresis in two dimensions in agarose gels. *Nucleic Acids Res.* 9(24):6805-12. PubMed PMID: 6278448; PubMed Central PMCID: PMC327643.

Pressman S, Bei Y, Carthew R. (2007) SnapShot: Posttranscriptional Gene Silencing. *Cell.* 130(3):570-570.e1. doi: 10.1016/j.cell.2007.07.042.

Rand KN (1996) Crystal violet can be used to visualize DNA bands during gel electrophoresis and to improve cloning efficiency. *Trends J Tech Tips Online* 1:23–24. doi: 10.1016/S1366-2120(07)70016-8.

RDocumentation (n.d.) vipor package *Retrieved April 25th, 2020 from* https://www.rdocumentation.org/packages/vipor/versions/0.4.4

Singh H, Saxena S. (2003) A class of shrinkage estimators for variance of a normal population. *Brazilian Journal of Probability and Statistics.* 17(1), 41-56. *Retrieved September 3, 2020, from* http://www.jstor.org/stable/43601022

Schnütgen F, Doerflinger N, Calléja C, Wendling O, Chambon P, Ghyselinck NB. (2003) A directional strategy for monitoring Cre-mediated recombination at the cellular level in the mouse. *Nat Biotechnol.* 21(5):562-5. PubMed PMID: 12665802.

Schnütgen F, De-Zolt S, Van Sloun P, Hollatz M, Floss T, Hansen J, Altschmied J, Seisenberger C, Ghyselinck NB, Ruiz P, Chambon P, Wurst W, von Melchner H. (2005) Genomewide production of multipurpose alleles for the functional analysis of the mouse genome. *Proc Natl Acad Sci U S A.* 102(20):7221-6. PubMed PMID: 15870191; PubMed Central PMCID: PMC1129123.

Shevchuk NA, Bryksin AV, Nusinovich YA, Cabello FC, Sutherland M, Ladisch S. (2004) Construction of long DNA molecules using long PCR-based fusion of several fragments simultaneously. *Nucleic Acids Res.* 32(2):e19. PubMed PMID: 14739232; PubMed Central PMCID: PMC373371.

Shivdasani RA. (2006) MicroRNAs: regulators of gene expression and cell differentiation. *Blood.* 108(12):3646-53. PubMed PMID: 16882713; PubMed Central PMCID: PMC1895474.

Strachan T, Read A. (2010) *Human Molecular Genetics* (4th ed.). New York, United States: Garland Science.

Szczepanski SM, Knight RT. (2014) Insights into human behavior from lesions to the prefrontal cortex. *Neuron.* Sep 3;83(5):1002-18. doi: 10.1016/j.neuron.2014.08.011. PubMed PMID: 25175878; PubMed Central PMCID: PMC4156912.

ThermoFisher Scientific (n.d.) SYBR Safe - DNA Gel Stain *Retrieved May 15th, 2019 from* https://www.thermofisher.com/fi/en/home/life-science/dna-rna-purification-analysis/nucleic-acid-gel-electrophoresis/dna-stains/sybr-safe.html

Tian X, Seluanov A, Gorbunova V. (2017) Molecular Mechanisms Determining Lifespan in Short- and Long-Lived Species. *Trends Endocrinol Metab.* 28(10):722-734. doi: 10.1016/j.tem.2017.07.004. PubMed PMID: 28888702; PubMed Central PMCID: PMC5679293.

Tirabassi R. (n.d.) How to Identify Supercoils, Nicks and Circles in Plasmid Preps *Retrieved June 3rd, 2019 from* https://bitesizebio.com/13524/how-to-identify-supercoils-nicks-and-circles-in-plasmid-preps/

Tosato M, Zamboni V, Ferrini A, Cesari M. (2007) The aging process and potential interventions to extend life expectancy. *Clin Interv Aging.* 2(3):401-12. PubMed PMID: 18044191; PubMed Central PMCID: PMC2685272.

Tushev G, Glock C, Heumüller M, Biever A, Jovanovic M, Schuman EM. (2018) Alternative 3' UTRs Modify the Localization, Regulatory Potential, Stability, and Plasticity of mRNAs in Neuronal Compartments. *Neuron.* 98(3):495-511.e6. doi: 10.1016/j.neuron.2018.03.030. PubMed PMID: 29656876.

Wellmer FW. (1998) Standard Deviation and Variance of the Mean. *In: Statistical Evaluations in Exploration for Mineral Deposits. Springer, Berlin, Heidelberg.* doi: 10.1007/978-3-642-60262-7_6

Wu ML, Freitas SS, Monteiro GA, Prazeres DM, Santos JA. (2019) Stabilization of naked and condensed plasmid DNA against degradation induced by ultrasounds and high-shear vortices. *Biotechnol Appl Biochem*. 53(Pt 4):237-246. doi: 10.1042/BA20080215

Zhao W, Siegel D, Biton A, Tonqueze OL, Zaitlen N, Ahituv N, Erle DJ. (2017) CRISPR-Cas9-mediated functional dissection of 3'-UTRs. *Nucleic Acids Res.* 45(18):10800-10810. doi: 10.1093/nar/gkx675. PubMed PMID: 28985357; PubMed Central PMCID: PMC5737544.

9. Appendices

9.1. Appendix A: Measurement day and 3′UTR type abbreviations and identities

Table 3 records measurement day abbreviations and Table 4 3′UTR abbreviations.

Table 3. Measurement day abbreviations, according to which measurement of which group (‘batch’) was or were measured in each day.

Day abbreviation	Date	Description
1-0	2017-08-10	Protocol test with only a few Batch 1 plasmids
1-1	2017-09-05	Batch 1 (full), Day 1
1-2	2017-09-27	Batch 1, Day 2
1-3	2018-09-03	Batch 1, Day 3
2-1	2018-09-11	Batch 2, Day 1
2-2/3-1	2018-09-17	{Batch 2, Day 2} and {Batch 3, Day 1} (same substrates used)
3-2	2018-10-10	Batch 3, Day 2
4-0	2018-11-19	Batch 4, Co-transfection molar ratio testing only

Table 4. 3′UTR abbreviations and their associated information, grouped by identity type. Sequence lengths use (+#) to separately indicate downstream (+SV40) and genomic DNA downstream of a gene's cleavage-and-polyadenylation site, and do not take into account all restriction site lengths; see Appendix C for the primer sequences used. Reference sequences are in the form of plasmid name, Appendix reference, NCBI accession number, or Ensembl Genome Browser stable identifier.

3′UTR abbreviation	Reference sequence	Sequence length (base pairs)	Source identity
Viral sequences			
(+SV40)	pGL4.73	256	simian vacuolating virus 40 3′UTR, the default 3′UTR in the Promega dual-luciferase reporter assay plasmids, used here as a control
+WPRE+hGH	(see Appendix B)	1170	human growth hormone 1 3′UTR, here preceded by the Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element
Flex-cassette testing			

+0lox-F(+SV40)	(see Appendix B), pGL4.73	297(+256)	bovine growth hormone 3′UTR (forward direction), followed by (+SV40)
+1lox-F(+SV40)	(see Appendix B), pGL4.73	483(+256)	bovine growth hormone 3′UTR (forward direction), with loxP and lox5171 site, followed by (+SV40)
+flex-F(+SV40)	(see Appendix B), pGL4.73	622(+256)	bovine growth hormone 3′UTR (forward direction), with loxP and lox5171 sites, followed by (+SV40)
+flex-F(no SV40)	(see Appendix B)	622	bovine growth hormone 3′UTR (forward direction), with loxP and lox5171 sites
+0lox-B(+SV40)	(see Appendix B), pGL4.73	297(+256)	bovine growth hormone 3′UTR (backward direction), followed by (+SV40)
+1lox-B(+SV40)	(see Appendix B), pGL4.73	483(+256)	bovine growth hormone 3′UTR (backward direction), with loxP and lox5171 site, followed by (+SV40)
+flex-B(+SV40)	(see Appendix B), pGL4.73	622(+256)	bovine growth hormone 3′UTR (backward direction), with loxP and lox5171 sites, followed by (+SV40)
+flex-B(no SV40)	(see Appendix B)	622	bovine growth hormone 3′UTR (backward direction), with loxP and lox5171 sites
+flex-B+mDNAJB6-3	(see Appendix B), (same as DNAJB6-3)	716	bovine growth hormone 3′UTR (backward direction), with loxP and lox5171 sites, followed by +mDNAJB6-3
Anti-aggregation			
+mDNAJB6-3	NM_001037941.3	89(+0)	mouse DnaJ heat shock protein family (Hsp40) member B6, isoform b 3′UTR
+mDNAJB6-2	NM_011847.4	671(+0)	mouse DnaJ heat shock protein family (Hsp40) member B6, isoform c 3′UTR
+mDNAJB6-1	NM_001037940.4	1365(+0)	mouse DnaJ heat shock protein family (Hsp40) member B6, isoform a 3′UTR

+mDNAJB6-1-extended	ENSMUST00000008733.14	1365+571	mouse DnaJ heat shock protein family (Hsp40) member B6, isoform a 3′UTR and the following 571 bp
+mDNAJC6	NM_198412.2	2226(+0)	mouse DnaJ heat shock protein family (Hsp40) member C6
+mHSPA1A	NM_010479.2	641(+0)	mouse heat shock protein 1A 3′UTR
+mMITF	NM_001113198.1	3183(+0)	mouse melanogenesis associated transcription factor 3′UTR
+mMME	NM_001289463.1	3381(+0)	mouse membrane metallo endopeptidase 3′UTR
+mPCSK1N	NM_013892.3	1307(+0)	mouse proprotein convertase subtilisin/kexin type 1 inhibitor 3′UTR
+mSCG5	NM_009162.3	488(+0)	mouse secretogranin V 3′UTR
Aggregation proteins			
+mAPP	NM_007471.3	895(+0)	mouse amyloid beta (A4) precursor protein 3′UTR
+mAPP-extended	ENSMUST00000005406.11	898+998	mouse amyloid beta (A4) precursor protein 3′UTR and the following 998 bp
+mITM2B	NM_008410.2	827(+0)	mouse integral membrane protein 2B 3′UTR (an aggregation protein according to Garringer et al., 2017)
+mSNCA	NM_001042451.2	556(+0)	mouse synuclein, alpha 3′UTR
+mSNCA-extended	ENSMUST00000114268.4	550+972	mouse synuclein, alpha 3′UTR and the following 972 bp
+mIAPP-extended	ENSMUST00000041993.2	473+969	mouse islet amyloid polypeptide 3′UTR and the following 969 bp
+mPRNP-extended	ENSMUST00000091288.12	1235+716	mouse prion protein 3′UTR and the following 716 bp
+mACTB-extended	ENSMUST00000100497.10	683+530	mouse actin, beta 3′UTR and the following 530 bp
+mSOD1-extended	ENSMUST00000023707.9	65+594	mouse superoxide dismutase 1, soluble 3′UTR and the following 594 bp

+mTRP53-extended	ENSMUST00000171247.7	564+523	mouse transformation related protein 53 3′UTR and the following 523bp
Neuron survival			
+mGDF11	NM_010272.1	560(+0)	mouse growth differentiation factor 11 3′UTR
+mGDNF+60bp(+SV40)	NM_010275.3, pGL4.73	2748+60 (+256)	mouse glial cell line derived neurotrophic factor 3′UTR, followed by (+SV40)
+mGDNF+60bp(no SV40)	NM_010275.3	2748+60	mouse glial cell line derived neurotrophic factor 3′UTR, followed by (+SV40)
+mPARK2-1	NM_016694.4	1713(+0)	mouse E3 ubiquitin-protein ligase parkin, isoform 1 3′UTR

9.2. Appendix B: Non-genomic template sequences

9.2.1. Flex-cassette template

Flex-cassette (reverse orientation) (cloned into pUC57)

- loxP site – Underlined
- lox5171 site – *Underlined italics*
- bovine growth hormone 3′UTR in inverse orientation – **Bold**
- AATAAA polyadenylation signal in inverse orientation - **Bold highlighted**

GGCCGCGAGCTAGATCTAGAATAACTTCGTATAGCATACATTATACGAAGTTATGGGTCGATGGTG
AGATCTGGACTAGAGGGTCGATGGTGATGCTTGG*ATAACTTCGTATAGTACACATTATACGAAGTT*
*AT*CGGATCCCAGTGTGGTGGTACTCGAGGTCGACTCTAGAGGATCGAGCCCCAGCTGGTTCTT**TCC**
GCCTCAGAAGCCATAGAGCCCACCGCATCCCCAGCATGCCTGCTATTGTCTTCCCAATCCTCCCCC
TTGCTGTCCTGCCCCACCCCACCCCCCAGAATAGAATGACACCTACTCAGACAATGCGATGCAATT
TCCTCATTTTATTAGGAAAGGACAGTGGGAGTGGCACCTTCCAGGGTCAAGGAAGGCACGGGGGAG
GGGCAAACAACAGATGGCTGGCAACTAGAAGGCACAGTCGAGGCTGATCAGCGAGCTCTAGAGAAT
TGATCCCGAATTCGATAACTTCGTATAATGTATGCTATACGAAGTTATGGGTCGATGGTGATGCAT
GGCAATTCGGGTCGATGGTGAACTTGGCCCTACTAA*ATAACTTCGTATAATGTGTACTATACGAAG*
*TTAT*AAGCTCTAGACTAGATAATCCTGCA

Orientations:
-->loxP>-->lox5171>--<bGHpA<--<loxP<--<lox5171<--

9.2.2. WPRE+hGH template

- WPRE sequence – Underlined
- hGH 3′UTR sequence – **Bold underlined**

GTACAAGTAAAGCGGCCGCACTCGAGATATCAAGCTTATCGATAATCAACCTCTGGATTACAAAAT
TTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTT
AATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTG
GTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTT
TGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGC
TTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGC
TCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCTTTCCTTGGCTGCT
CGCCTATGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCC
AGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCC
TCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCATCGATACCGAGCGCTGCTCGAGAGATC
TACGGGTGGCATCCCTGTGACCCCTCCCCAGTGCCTCTCCTGGCCCTGGAAGTTGCCACTCCAGTG
CCCACCAGCCTTGTCCTAATAAAATTAAGTTGCATCATTTTGTCTGACTAGGTGTCCTTCTATAAT
ATTATGGGGTGGAGGGGGGTGGTATGGAGCAAGGGGCAAGTTGGGAAGACAA**CCTGTAGGGCCTGC**
GGGGTCTATTGGGAACCAAGCTGGAGTGCAGTGGCACAATCTTGGCTCACTGCAATCTCCGCCTCC
TGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCCGAGTTGTTGGGATTCCAGGCATGCATGACCAGG
CTCAGCTAATTTTTGTTTTTTTGGTAGAGACGGGGTTTCACCATATTGGCCAGGCTGGTCTCCAAC
TCCTAATCTCAGGTGATCTACCCACCTTGGCCTCCCAAATTGCTGGGATTACAGGCGTGAACCACT
GCTCCCTTCCCTGTCCTTCTGATTTTGTAGGTAACCACGTGCGGACCG

9.3. Appendix C: Primers used

Table 5 records all primers used in this project.

Table 5. Sequencing and PCR primers used for this thesis's plasmids. Upper-case letters refer to annealing sequences while lower-case letters refer to added restriction sites and their AT-rich 6-nucleotide leading sequences. Restriction sites are indicated in the names where present, and F and R are used for Forward and Reverse-complement primers respectively; the no_SV40 primers are ordered with the R first because their PCR product is everything outside the SV40 3′UTR, with the two AscI restriction sites ligated to make a circular plasmid.

Sequencing primers	
pGL4.73_seq_F	ACATCAAGAGCTTCGTGGAG
pGL4.73_seq_R (anneals within (+SV40))	CAAACTCATCAATGTATCTTATC
pGL4.73_seq_R2 (anneals after SalI site)	GAAGACAGTCATAAGTGCGG
PCR primers	
Viral sequences	
WPRE_NheI_F	aataatgctagcGTACAAGTAAAGCGGCCGCAC
WPRE_bGHpA_SalI_R	aataatgtcgaCGGTCCGCACGTGGTTAC
no-SV40_AscI_R	attattggcgcGCCGCCCCGACTC
no-SV40_AscI_F	attattggcgcgccGATAAGGATCCGTCGACC
Flex-cassette testing	
bGHpA_0lox_MluI_NheI_F	atattagctagcacgcgtCGAATTCGGGATCAATTCTC
bGHpA_0lox_NheI_R	ttattagctaGCCCCAGCTGGTTC
bGHpA_1lox_MluI_NheI_F	atattagctagcacgcgTTTAGTAGGGCCAAGTTCAC

bGHpA_1lox_NheI_R	ttattagctagcGTCGATGGTGATGCTTG
bGHpA_flex_MluI_NheI_F	atattagctagcacgcgTGCAGGATTATCTAGTCTAGAGCTTATAAC
bGHpA_flex_NheI_R	ttattagctaGCCGCGAGCTAGATCTAGAA
mDNAJB6-3_MluI_F	aataatacgcgtACAAAAAGGAATGGCCTGG
mDNAJB6-3_SaII_R	aataatgtcgacGGTGATAGAAGCAATTCTTTTTATTCTAG
Anti-aggregation	
mDNAJB6-3_NheI_F	aataatgctagcACAAAAAGGAATGGCCTGG
mDNAJB6-3_SaII_R	aataatgtcgacGGTGATAGAAGCAATTCTTTTTATTCTAG
mDNAJB6-2_OuterF	TTACAGGATTGTGGAGAACGGT
mDNAJB6-2_OuterR	AGCACTAATGAGGTGACACTTAGA
mDNAJB6-2_XbaI_InnerF	taataatctagaCTCAACGCACGCATTTAAC
mDNAJB6-2_SalI_InnerR	attattgtcgacAGACCCCACTCATGAAAGA
mDNAJB6-1_XbaI_F	aataattctagaGCTGGACTTGGCG
mDNAJB6-1_SalI_R	aataatgtcgacAAATGTATAAAACTGTATTTAAAAAAAC
mDNAJB6-1-extended_SalI_R	aataatgtcgacTAGTGTTAGGGTTAAGAGT
mDNAJC6_XbaI_F	ttattatctagaTTTGTGAGCTTTTCTATGCTG
mDNAJC6_SalI_R	ttctatgtcgacTATTTTGTATCCTTTTTAATATACATTAC
mHSPA1A_XbaI_F	tataattctagAGGCCTCTGCTGGCTCTCCCGG
mHSPA1A_SalI_R	aataatgtcgacTGGCAAGTGATTTTGAAGTTTATTTTCAAATGACCGG
mMITF_OuterF	ATGCATTTGGGTAACCGCAC
mMITF_OuterR	GTGCCATCTTCGAATGACACAAG
mMITF_NheI_InnerF	attattgctagCGAGCCTGCCTTGC
mMITF_SalI_InnerR	taataagtcgacAACACAATGTGAAAAACCAAATG
mMME_XbaI_F	aataattctagaTCTTCACAAGATACTGAACATCC
mMME_SalI_R	ttattagtcgacTTGACACTAAGAAAATATAATTAGAGAAAAAG
mPCSK1N_XbaI_F	aataattctagaGCGCTGCTGCATCCTG
mPCSK1N_SalI_R	aataatgtcgacTTACCACAGACATCTTTTTATTGTCATTTG
mSCG5_XbaI_F	aataattctAGAGAAGACAGTATGTAGAAACC
mSCG5_SalI_R	ttattagtcgacATTAAGCAAAGAAATTGATTTTATAC
Aggregation proteins	
mAPP_XbaI_F	ttattatctagaGCCCCACCCGCGC
mAPP_SalI_R	ttattagtcgacTCTTGCCTGGAGTTATTTTATTTAATTTATTTATGTAATACAG
mAPP_NheI_F	ttattagctagcGCCCCACCCGCGC
mAPP-extended_SalI_R	ttattagtcgacTCAGGAGGATAAGGCCATCTTTG
mITM2B_XbaI_F	ttattatctaGAAGTCAAGAAAAAACGTGG
mITM2B_SalI_R	ttattagtcgacTTTAGTAATTAATGGCTTATTTTATTTTTTTAAAAAG
mSNCA_XbaI_F	ttattatctaGAATGTCATTGCACCCAATC
mSNCA_SalI_R	aataatgtcgacGATTTTGCAATGGATAATATTTTATT
mSNCA-extended_SalI_R	ttattagtcgaCTAAACACCATCTGGGCTAC
mIAPP_OuterF	CCAACCAACGTGGGATCGAA
mIAPP_OuterR	TTTTCAGGCACTGCTCCGT
mIAPP_NheI_InnerF	aattaagctagcAGTCAATGTACTTCTGCAGCACTTAATAC

mIAPP-extended _SalI_InnerR	aataatgtcgaCCATTCTTCCGTGACTGAACC
mPRNP_OuterF	AGGACCGCTACTACCGTGAA
mPRNP_OuterR	GCGCTATCACGTTGTGGTTT
mPRNP_BcuI_InnerF	aataatactagtGGGAGGCCTTCCTGCTTG
mPRNP-extended _SalI_InnerR	aataatgtcgacTGTCTGAGCCAGGTATGGGA
mACTB_OuterF	CATCACTATTGGCAACGAGCG
mACTB_OuterR	GGTAGCTTATTTAGGGTTGGC
mACTB_XbaI_InnerF	aataattctagaGCGGACTGTTACTGAGCTGC
mACTB-extended _SalI_InnerR	aataatgtcgaCAAAGCTAACTTGCGGCACC
mSOD1_OuterF	GAGCCGTCTTCCCAAGTTACC
mSOD1_OuterR	CAGCTTTCTGGTGACCCTACTTA
mSOD1_XbaI_InnerF	aataattctagACATTCCCTGTGTGGTCTGAGTC
mSOD1-extended _SalI_InnerR	aataatgtcgACTTGGTGTCATTCCCCGAG
mTRP53_OuterF	CGAGATGTTCCGGGAGCTGA
mTRP53_OuterR	TAGTTCTCAGTAAAGCTGTGGGAC
mTRP53_XbaI_InnerF	aataattctagaCTCCCATCACTTCATCCCTCC
mTRP53-extended _SalI_InnerR	aataatgtcgacGAGTTTGGGGATGCTGGGAA
Neuron survival	
mGDF11_OuterF	CCTGGACTGCGATGAACACT
mGDF11_OuterR	AAGGTAGGCCTGAGGGGTAG
mGDF11_XbaI_InnerF	taataatctagaGTTGTGGGCTACAGTGGAT
mGDF11_SalI_InnerR	attattgtcgacTCCTTACTTTGCCCCATC
mGDNF_NheI_F (an attempt to recreate the +mGDNF+60bp(+SV40) construct)	aattaagctagCCCCGGCTCCAGAGACTG
mGDNF_NheI_R (an attempt to recreate the +mGDNF+60bp(+SV40) construct)	aattaagctaGCCTGCCCGGCCAAGGC
mPARK2-1_XbaI_F	aataattctAGAGAGATGTCACTTGGCCCTG
mPARK2-1_SalI_R	attgatgtcgacGATTCTTTCATTGACAGTCTGGGTC

9.4. Appendix D: Raw and processed measurement data

Tables 6 and 7 record the raw and processed luminometer measurements respectively, split into two tables to allow for page width.

Firefly luciferase was expressed by the internal control plasmid pGL4.73 and Renilla luciferase by each reporter plasmid derived from plasmid pGL4.73. The control:reporter plasmid co-transfection

ratio is 1:11.8 for days 1-0, 1-1, and 1-2, and 1:10 for all other days except where indicated with a different ratio in the 3′UTR designation.

The 3′UTR designation also indicates number of measurements for each 3′UTR type in the form '_(n=#)'.

The day and 3′UTR type abbreviations are described in Appendix A.

'Measurement' refers to the measurement number within each replicate within a given measurement day.

Table 6. Luminometer unchanged measurements, as described above.

3′UTR	Batch	Day	Measurement	Firefly luciferase	Renilla luciferase
Viral sequences					
(+SV40)_(n=24)	1	1-0	1	5,308,965	95,362,784
(+SV40)_(n=24)	1	1-0	2	5,852,236	105,678,000
(+SV40)_(n=24)	1	1-0	3	5,651,853	102,565,264
(+SV40)_(n=24)	1	1-1	1	2,356,662	44,999,892
(+SV40)_(n=24)	1	1-1	2	3,128,744	80,371,248
(+SV40)_(n=24)	1	1-1	3	1,815,211	47,445,252
(+SV40)_(n=24)	1	1-2	1	1,701,942	50,939,964
(+SV40)_(n=24)	1	1-2	2	3,130,801	119,524,824
(+SV40)_(n=24)	1	1-2	3	2,828,680	99,335,328
(+SV40)_(n=24)	1	1-3	2	926,643	51,865,592
(+SV40)_(n=24)	1	1-3	3	1,008,143	55,199,848
(+SV40)_(n=24)	2	2-1	1	663,556	66,007,720
(+SV40)_(n=24)	2	2-1	2	1,388,227	148,422,688
(+SV40)_(n=24)	2	2-1	3	1,280,164	134,326,704
(+SV40)_(n=24)	2	2-2/3-1	1	3,457,175	176,843,200
(+SV40)_(n=24)	2	2-2/3-1	2	3,163,427	164,197,696
(+SV40)_(n=24)	2	2-2/3-1	3	3,498,232	172,149,728
(+SV40)_(n=24)	3	2-2/3-1	1	8,886,113	555,942,592
(+SV40)_(n=24)	3	2-2/3-1	2	9,437,454	552,014,464
(+SV40)_(n=24)	3	2-2/3-1	3	8,824,526	527,984,768
(+SV40)_(n=24)	3	3-2	1	2,067,841	287,839,296
(+SV40)_(n=24)	3	3-2	2	2,219,196	342,387,616
(+SV40)_(n=24)	3	3-2	3	2,488,922	319,845,696
+WPRE+hGH_(n=6)	3	2-2/3-1	1	3,669,220	963,177,472
+WPRE+hGH_(n=6)	3	2-2/3-1	2	4,150,225	1,074,323,712
+WPRE+hGH_(n=6)	3	2-2/3-1	3	4,580,382	1,121,431,168
+WPRE+hGH_(n=6)	3	3-2	1	1,535,947	621,207,744
+WPRE+hGH_(n=6)	3	3-2	2	1,076,537	497,969,920
+WPRE+hGH_(n=6)	3	3-2	3	1,352,831	580,466,560
Flex-cassette testing					
+0lox-F(+SV40)_(n=12)	2	2-1	1	965,991	36,619,148
+0lox-F(+SV40)_(n=12)	2	2-1	2	1,158,535	42,851,960
+0lox-F(+SV40)_(n=12)	2	2-1	3	1,153,313	46,132,972
+0lox-F(+SV40)_(n=12)	2	2-2/3-1	1	4,030,095	98,602,368
+0lox-F(+SV40)_(n=12)	2	2-2/3-1	2	3,820,136	92,704,928
+0lox-F(+SV40)_(n=12)	2	2-2/3-1	3	3,621,798	83,684,016

+0lox-F(+SV40)_(n=12)	3	2-2/3-1	1	7,524,540	159,939,328
+0lox-F(+SV40)_(n=12)	3	2-2/3-1	2	8,557,831	176,001,296
+0lox-F(+SV40)_(n=12)	3	3-2	1	1,983,213	96,376,640
+0lox-F(+SV40)_(n=12)	3	3-2	2	2,827,349	133,969,264
+0lox-F(+SV40)_(n=12)	3	3-2	3	2,995,375	134,752,144
+1lox-F(+SV40)_(n=6)	2	2-1	1	1,032,376	30,432,402
+1lox-F(+SV40)_(n=6)	2	2-1	2	1,175,686	30,363,526
+1lox-F(+SV40)_(n=6)	2	2-1	3	1,326,094	34,693,364
+1lox-F(+SV40)_(n=6)	2	2-2/3-1	1	4,601,312	79,110,384
+1lox-F(+SV40)_(n=6)	2	2-2/3-1	2	4,433,532	75,665,328
+1lox-F(+SV40)_(n=6)	2	2-2/3-1	3	3,936,512	68,866,784
+flex-F(+SV40)_(n=18)	1	1-0	1	237,688	5,007,632
+flex-F(+SV40)_(n=18)	1	1-0	2	232,388	5,619,819
+flex-F(+SV40)_(n=18)	1	1-0	3	257,786	5,265,715
+flex-F(+SV40)_(n=18)	1	1-1	1	4,553,205	19,705,838
+flex-F(+SV40)_(n=18)	1	1-1	2	6,107,946	28,253,350
+flex-F(+SV40)_(n=18)	1	1-1	3	6,381,508	27,081,802
+flex-F(+SV40)_(n=18)	1	1-2	1	7,584,637	57,428,700
+flex-F(+SV40)_(n=18)	1	1-2	2	5,536,485	45,111,844
+flex-F(+SV40)_(n=18)	1	1-2	3	7,073,510	54,040,556
+flex-F(+SV40)_(n=18)	1	1-3	1	877,871	14,794,245
+flex-F(+SV40)_(n=18)	1	1-3	2	1,079,873	16,717,230
+flex-F(+SV40)_(n=18)	1	1-3	3	1,039,361	16,988,414
+flex-F(+SV40)_(n=18)	2	2-1	1	840,512	19,415,054
+flex-F(+SV40)_(n=18)	2	2-1	2	1,025,992	22,913,550
+flex-F(+SV40)_(n=18)	2	2-1	3	916,677	16,720,136
+flex-F(+SV40)_(n=18)	2	2-2/3-1	1	3,493,269	46,324,892
+flex-F(+SV40)_(n=18)	2	2-2/3-1	2	4,015,976	55,527,936
+flex-F(+SV40)_(n=18)	2	2-2/3-1	3	3,519,718	50,399,928
+flex-F (no_SV40)_(n=6)	2	2-1	1	613,516	12,881,890
+flex-F (no_SV40)_(n=6)	2	2-1	2	729,500	14,505,879
+flex-F (no_SV40)_(n=6)	2	2-1	3	867,171	17,641,156
+flex-F (no_SV40)_(n=6)	2	2-2/3-1	1	2,365,682	28,609,076
+flex-F (no_SV40)_(n=6)	2	2-2/3-1	2	2,526,322	28,057,574
+flex-F (no_SV40)_(n=6)	2	2-2/3-1	3	2,119,628	24,088,124
+0lox-B(+SV40)_(n=6)	2	2-1	1	771,139	66,746,668
+0lox-B(+SV40)_(n=6)	2	2-1	2	993,762	78,604,496
+0lox-B(+SV40)_(n=6)	2	2-1	3	1,130,521	96,417,240
+0lox-B(+SV40)_(n=6)	2	2-2/3-1	1	3,430,323	179,917,040
+0lox-B(+SV40)_(n=6)	2	2-2/3-1	2	3,541,597	191,829,168
+0lox-B(+SV40)_(n=6)	2	2-2/3-1	3	3,337,722	169,485,168
+1lox-B(+SV40)_(n=6)	2	2-1	1	1,070,179	32,391,502
+1lox-B(+SV40)_(n=6)	2	2-1	2	852,118	24,513,854

+1lox-B(+SV40)_(n=6)	2	2-1	3	1,022,381	27,928,394
+1lox-B(+SV40)_(n=6)	2	2-2/3-1	1	7,073,542	126,191,720
+1lox-B(+SV40)_(n=6)	2	2-2/3-1	2	7,087,911	134,069,712
+1lox-B(+SV40)_(n=6)	2	2-2/3-1	3	6,391,504	127,111,416
+flex-B(+SV40)_(n=18)	1	1-0	1	4,919,098	10,511,795
+flex-B(+SV40)_(n=18)	1	1-0	2	4,964,934	9,748,039
+flex-B(+SV40)_(n=18)	1	1-0	3	4,673,697	9,980,481
+flex-B(+SV40)_(n=18)	1	1-1	1	4,036,619	11,264,284
+flex-B(+SV40)_(n=18)	1	1-1	2	4,801,505	12,712,927
+flex-B(+SV40)_(n=18)	1	1-1	3	3,274,777	5,770,324
+flex-B(+SV40)_(n=18)	1	1-2	1	7,353,360	35,624,556
+flex-B(+SV40)_(n=18)	1	1-2	2	5,212,015	28,162,058
+flex-B(+SV40)_(n=18)	1	1-2	3	7,267,230	39,599,704
+flex-B(+SV40)_(n=18)	1	1-3	1	837,932	5,004,004
+flex-B(+SV40)_(n=18)	1	1-3	2	885,538	5,316,198
+flex-B(+SV40)_(n=18)	1	1-3	3	771,007	5,489,492
+flex-B(+SV40)_(n=18)	2	2-1	1	441,389	5,430,486
+flex-B(+SV40)_(n=18)	2	2-1	2	1,019,031	13,526,798
+flex-B(+SV40)_(n=18)	2	2-1	3	955,828	12,355,942
+flex-B(+SV40)_(n=18)	2	2-2/3-1	1	3,889,764	36,137,300
+flex-B(+SV40)_(n=18)	2	2-2/3-1	2	4,436,120	41,024,220
+flex-B(+SV40)_(n=18)	2	2-2/3-1	3	4,056,054	34,060,836
+flex-B (no_SV40)_(n=6)	2	2-1	1	880,624	1,439,577
+flex-B (no_SV40)_(n=6)	2	2-1	2	950,408	1,522,968
+flex-B (no_SV40)_(n=6)	2	2-1	3	930,524	1,491,788
+flex-B (no_SV40)_(n=6)	2	2-2/3-1	1	2,065,421	1,608,606
+flex-B (no_SV40)_(n=6)	2	2-2/3-1	2	2,272,481	1,621,328
+flex-B (no_SV40)_(n=6)	2	2-2/3-1	3	1,718,645	1,363,163
+flex-B+mDNAJB6-3_(n=9)	1	1-2	1	6,167,922	1,618,802
+flex-B+mDNAJB6-3_(n=9)	1	1-2	2	5,626,580	1,380,764
+flex-B+mDNAJB6-3_(n=9)	1	1-2	3	5,593,556	1,302,744
+flex-B+mDNAJB6-3_(n=9)	1	1-3	1	776,037	700,177
+flex-B+mDNAJB6-3_(n=9)	1	1-3	2	874,875	857,030
+flex-B+mDNAJB6-3_(n=9)	1	1-3	3	801,664	662,533
+flex-B+mDNAJB6-3_(n=9)	3	2-2/3-1	1	3,346,106	2,283,794

+flex-B+mDNAJB6-3_(n=9)	3	2-2/3-1	2	7,232,244	4,814,509
+flex-B+mDNAJB6-3_(n=9)	3	2-2/3-1	3	7,891,228	5,118,635
Anti-aggregation					
+mDNAJB6-3_(n=15)	1	1-1	1	2,184,594	23,967,620
+mDNAJB6-3_(n=15)	1	1-1	2	2,485,570	28,804,618
+mDNAJB6-3_(n=15)	1	1-1	3	2,894,751	34,286,884
+mDNAJB6-3_(n=15)	1	1-2	1	1,394,323	15,262,985
+mDNAJB6-3_(n=15)	1	1-2	2	2,198,623	26,113,408
+mDNAJB6-3_(n=15)	1	1-2	3	1,863,264	23,501,858
+mDNAJB6-3_(n=15)	1	1-3	1	425,710	3,218,790
+mDNAJB6-3_(n=15)	1	1-3	2	535,466	4,040,734
+mDNAJB6-3_(n=15)	1	1-3	3	683,633	5,178,380
+mDNAJB6-3_(n=15)	2	2-1	1	462,698	5,647,587
+mDNAJB6-3_(n=15)	2	2-1	2	689,911	9,351,981
+mDNAJB6-3_(n=15)	2	2-1	3	642,140	8,676,932
+mDNAJB6-3_(n=15)	2	2-2/3-1	1	1,928,097	14,048,780
+mDNAJB6-3_(n=15)	2	2-2/3-1	2	2,170,199	15,575,222
+mDNAJB6-3_(n=15)	2	2-2/3-1	3	1,588,989	11,967,514
+mDNAJB6-2_(n=6)	2	2-1	1	1,049,891	157,114,432
+mDNAJB6-2_(n=6)	2	2-1	2	1,146,044	180,627,872
+mDNAJB6-2_(n=6)	2	2-1	3	990,731	163,446,288
+mDNAJB6-2_(n=6)	2	2-2/3-1	1	1,551,631	114,651,864
+mDNAJB6-2_(n=6)	2	2-2/3-1	2	1,760,207	137,206,688
+mDNAJB6-2_(n=6)	2	2-2/3-1	3	1,519,321	122,818,040
+mDNAJB6-1_(n=21)	1	1-1	1	3,158,073	3,078,380
+mDNAJB6-1_(n=21)	1	1-1	2	4,431,315	4,354,015
+mDNAJB6-1_(n=21)	1	1-1	3	3,651,141	3,786,360
+mDNAJB6-1_(n=21)	1	1-2	1	4,897,150	5,163,375
+mDNAJB6-1_(n=21)	1	1-2	2	3,471,893	3,463,889
+mDNAJB6-1_(n=21)	1	1-2	3	4,722,548	5,377,855
+mDNAJB6-1_(n=21)	1	1-3	1	739,151	2,087,518
+mDNAJB6-1_(n=21)	1	1-3	2	914,426	2,208,191
+mDNAJB6-1_(n=21)	1	1-3	3	978,536	2,407,067
+mDNAJB6-1_(n=21)	2	2-1	1	649,771	2,153,218
+mDNAJB6-1_(n=21)	2	2-1	2	901,675	3,034,103
+mDNAJB6-1_(n=21)	2	2-1	3	1,101,410	3,811,986
+mDNAJB6-1_(n=21)	2	2-2/3-1	1	4,538,213	11,608,278
+mDNAJB6-1_(n=21)	2	2-2/3-1	2	4,957,758	12,209,776
+mDNAJB6-1_(n=21)	2	2-2/3-1	3	4,761,719	11,257,890
+mDNAJB6-1_(n=21)	3	2-2/3-1	1	8,228,159	21,523,228
+mDNAJB6-1_(n=21)	3	2-2/3-1	2	10,729,239	27,801,124
+mDNAJB6-1_(n=21)	3	2-2/3-1	3	8,798,327	22,804,666
+mDNAJB6-1_(n=21)	3	3-2	1	1,730,446	13,076,750
+mDNAJB6-1_(n=21)	3	3-2	2	2,734,932	18,451,030
+mDNAJB6-1_(n=21)	3	3-2	3	3,129,602	20,501,502
+mDNAJB6-1-extended_(n=6)	3	2-2/3-1	1	9,773,793	19,756,516

+mDNAJB6-1-extended_(n=6)	3	2-2/3-1	2	9,754,878	18,297,524
+mDNAJB6-1-extended_(n=6)	3	2-2/3-1	3	11,199,219	19,425,974
+mDNAJB6-1-extended_(n=6)	3	3-2	1	3,617,104	17,996,350
+mDNAJB6-1-extended_(n=6)	3	3-2	2	3,189,414	16,357,744
+mDNAJB6-1-extended_(n=6)	3	3-2	3	3,299,528	15,360,104
+mDNAJC6_(n=9)	1	1-1	1	1,931,648	158,827,584
+mDNAJC6_(n=9)	1	1-1	2	1,396,543	112,342,888
+mDNAJC6_(n=9)	1	1-1	3	1,309,534	99,322,608
+mDNAJC6_(n=9)	1	1-2	1	1,370,722	113,855,592
+mDNAJC6_(n=9)	1	1-2	2	1,580,304	149,204,896
+mDNAJC6_(n=9)	1	1-2	3	1,549,482	97,951,640
+mDNAJC6_(n=9)	1	1-3	1	583,241	54,431,056
+mDNAJC6_(n=9)	1	1-3	2	744,495	48,563,652
+mDNAJC6_(n=9)	1	1-3	3	626,587	54,272,068
+mHSPA1A_(n=12)	1	1-0	1	5,141,962	54,533,684
+mHSPA1A_(n=12)	1	1-0	2	5,302,622	61,757,752
+mHSPA1A_(n=12)	1	1-0	3	4,336,639	49,957,260
+mHSPA1A_(n=12)	1	1-1	1	1,440,472	23,083,028
+mHSPA1A_(n=12)	1	1-1	2	2,496,694	46,640,072
+mHSPA1A_(n=12)	1	1-1	3	2,836,981	57,286,184
+mHSPA1A_(n=12)	1	1-2	1	3,388,763	93,405,176
+mHSPA1A_(n=12)	1	1-2	2	3,418,954	97,135,920
+mHSPA1A_(n=12)	1	1-2	3	2,860,218	79,377,208
+mHSPA1A_(n=12)	1	1-3	1	709,958	20,122,718
+mHSPA1A_(n=12)	1	1-3	2	910,646	22,891,532
+mHSPA1A_(n=12)	1	1-3	3	882,158	24,398,922
+mMITF_(n=6)	2	2-1	1	791,040	26,788,976
+mMITF_(n=6)	2	2-1	2	923,111	36,661,504
+mMITF_(n=6)	2	2-1	3	1,007,568	42,032,740
+mMITF_(n=6)	2	2-2/3-1	1	6,209,355	190,394,480
+mMITF_(n=6)	2	2-2/3-1	2	5,899,692	190,770,384
+mMITF_(n=6)	2	2-2/3-1	3	5,867,711	184,840,304
+mMME_(n=9)	1	1-1	1	3,007,419	12,700,351
+mMME_(n=9)	1	1-1	2	3,134,802	14,563,416
+mMME_(n=9)	1	1-1	3	4,235,676	18,142,974
+mMME_(n=9)	1	1-2	1	2,061,369	12,877,383
+mMME_(n=9)	1	1-2	2	1,073,664	5,433,800
+mMME_(n=9)	1	1-2	3	1,291,699	7,724,833
+mMME_(n=9)	1	1-3	1	594,174	2,585,383
+mMME_(n=9)	1	1-3	2	664,931	2,965,592
+mMME_(n=9)	1	1-3	3	787,140	3,470,179
+mPCSK1N_(n=9)	1	1-1	1	3,140,344	56,651,664
+mPCSK1N_(n=9)	1	1-1	2	3,677,602	76,443,296
+mPCSK1N_(n=9)	1	1-2	1	4,035,710	126,664,632

+mPCSK1N_(n=9)	1	1-2	2	3,282,274	102,504,520
+mPCSK1N_(n=9)	1	1-2	3	4,047,297	127,926,408
+mPCSK1N_(n=9)	1	1-3	1	770,921	19,754,864
+mPCSK1N_(n=9)	1	1-3	2	1,174,415	31,148,554
+mPCSK1N_(n=9)	1	1-3	3	1,353,450	33,580,592
+mSCG5_(n=9)	1	1-1	1	2,995,877	30,688,960
+mSCG5_(n=9)	1	1-1	2	3,306,599	36,834,836
+mSCG5_(n=9)	1	1-1	3	2,591,768	26,755,616
+mSCG5_(n=9)	1	1-2	1	2,813,309	61,806,740
+mSCG5_(n=9)	1	1-2	2	3,824,876	74,305,968
+mSCG5_(n=9)	1	1-2	3	3,086,331	62,317,556
+mSCG5_(n=9)	1	1-3	1	757,158	21,024,166
+mSCG5_(n=9)	1	1-3	2	806,551	23,065,666
+mSCG5_(n=9)	1	1-3	3	716,949	20,741,116
Aggregation proteins					
+mAPP_(n=18)	1	1-0	1	3,764,489	193,570,528
+mAPP_(n=18)	1	1-0	2	3,451,815	188,737,152
+mAPP_(n=18)	1	1-0	3	3,538,620	179,744,464
+mAPP_(n=18)	1	1-1	1	2,709,339	169,222,704
+mAPP_(n=18)	1	1-1	2	2,423,342	169,018,752
+mAPP_(n=18)	1	1-1	3	3,476,245	227,390,240
+mAPP_(n=18)	1	1-2	1	3,036,727	326,105,824
+mAPP_(n=18)	1	1-2	2	4,198,443	486,385,888
+mAPP_(n=18)	1	1-2	3	3,638,537	436,705,728
+mAPP_(n=18)	1	1-3	1	433,188	38,325,884
+mAPP_(n=18)	1	1-3	2	625,248	56,626,980
+mAPP_(n=18)	1	1-3	3	570,287	49,791,812
+mAPP_(n=18)	3	2-2/3-1	1	6,445,421	521,362,240
+mAPP_(n=18)	3	2-2/3-1	2	6,380,531	516,765,120
+mAPP_(n=18)	3	3-2	1	1,926,196	321,567,040
+mAPP_(n=18)	3	3-2	2	1,792,884	347,593,024
+mAPP_(n=18)	3	3-2	3	2,215,148	343,203,712
+mAPP-extended_(n=6)	3	2-2/3-1	1	7,405,352	581,122,752
+mAPP-extended_(n=6)	3	2-2/3-1	2	7,837,914	614,488,896
+mAPP-extended_(n=6)	3	2-2/3-1	3	8,975,949	748,125,440
+mAPP-extended_(n=6)	3	3-2	1	2,119,866	362,173,216
+mAPP-extended_(n=6)	3	3-2	2	1,865,424	285,733,280
+mAPP-extended_(n=6)	3	3-2	3	2,511,367	390,114,624
+mITM2B_(n=12)	1	1-0	1	3,298,844	37,082,548
+mITM2B_(n=12)	1	1-0	2	4,256,839	46,073,044
+mITM2B_(n=12)	1	1-0	3	4,327,003	43,910,656
+mITM2B_(n=12)	1	1-1	1	2,635,957	36,343,844
+mITM2B_(n=12)	1	1-1	2	3,552,630	58,466,496
+mITM2B_(n=12)	1	1-1	3	1,914,405	36,919,696
+mITM2B_(n=12)	1	1-2	1	5,155,957	85,657,896
+mITM2B_(n=12)	1	1-2	2	8,666,783	131,203,864
+mITM2B_(n=12)	1	1-2	3	5,178,267	85,585,912
+mITM2B_(n=12)	1	1-3	1	445,018	9,593,262

+mITM2B_(n=12)	1	1-3	2	531,967	13,307,362
+mITM2B_(n=12)	1	1-3	3	689,067	15,538,999
+mSNCA_(n=15)	1	1-1	1	2,198,095	69,516,192
+mSNCA_(n=15)	1	1-1	2	2,409,681	99,285,032
+mSNCA_(n=15)	1	1-1	3	3,081,830	129,965,384
+mSNCA_(n=15)	1	1-2	1	4,013,221	222,530,432
+mSNCA_(n=15)	1	1-2	2	4,386,535	241,048,320
+mSNCA_(n=15)	1	1-2	3	3,697,762	201,459,888
+mSNCA_(n=15)	1	1-3	1	790,871	39,950,396
+mSNCA_(n=15)	1	1-3	2	716,877	37,297,904
+mSNCA_(n=15)	1	1-3	3	579,917	38,592,984
+mSNCA_(n=15)	3	2-2/3-1	1	4,296,286	226,028,352
+mSNCA_(n=15)	3	2-2/3-1	2	5,745,563	311,128,768
+mSNCA_(n=15)	3	2-2/3-1	3	4,926,860	260,769,088
+mSNCA_(n=15)	3	3-2	1	42,866	3,059,793
+mSNCA_(n=15)	3	3-2	2	186,036	15,524,640
+mSNCA_(n=15)	3	3-2	3	157,891	12,006,691
+mSNCA-extended_(n=6)	3	2-2/3-1	1	4,752,943	175,525,696
+mSNCA-extended_(n=6)	3	2-2/3-1	2	4,770,641	176,043,024
+mSNCA-extended_(n=6)	3	2-2/3-1	3	4,396,205	161,576,736
+mSNCA-extended_(n=6)	3	3-2	1	1,850,154	119,695,352
+mSNCA-extended_(n=6)	3	3-2	2	2,078,214	149,039,184
+mSNCA-extended_(n=6)	3	3-2	3	1,814,852	117,684,968
+mIAPP-extended_(n=6)	3	2-2/3-1	1	7,055,182	200,397,232
+mIAPP-extended_(n=6)	3	2-2/3-1	2	7,886,051	216,894,272
+mIAPP-extended_(n=6)	3	2-2/3-1	3	8,202,903	243,098,128
+mIAPP-extended_(n=6)	3	3-2	2	2,067,639	122,373,624
+mIAPP-extended_(n=6)	3	3-2	3	1,982,986	107,966,032
+mPRNP-extended_(n=6)	3	2-2/3-1	2	10,432,471	68,387,624
+mPRNP-extended_(n=6)	3	2-2/3-1	3	11,316,850	75,078,040
+mPRNP-extended_(n=6)	3	3-2	1	3,508,529	45,712,464
+mPRNP-extended_(n=6)	3	3-2	2	3,510,653	45,031,552
+mPRNP-extended_(n=6)	3	3-2	3	3,137,964	44,469,736

+mACTB-extended_(n=6)	3	2-2/3-1	2	10,003,239	896,003,072
+mACTB-extended_(n=6)	3	2-2/3-1	3	10,406,100	992,234,560
+mACTB-extended_(n=6)	3	3-2	1	3,122,085	540,008,960
+mACTB-extended_(n=6)	3	3-2	2	2,944,685	499,472,224
+mACTB-extended_(n=6)	3	3-2	3	2,694,632	471,478,784
+mSOD1-extended_(n=6)	3	2-2/3-1	1	9,958,320	129,519,184
+mSOD1-extended_(n=6)	3	2-2/3-1	2	9,455,389	128,769,544
+mSOD1-extended_(n=6)	3	2-2/3-1	3	9,634,607	125,221,240
+mSOD1-extended_(n=6)	3	3-2	1	2,943,328	88,220,248
+mSOD1-extended_(n=6)	3	3-2	2	2,882,976	81,957,096
+mSOD1-extended_(n=6)	3	3-2	3	2,612,252	76,062,232
+mTRP53-extended_(n=6)	3	2-2/3-1	1	5,664,725	55,817,852
+mTRP53-extended_(n=6)	3	2-2/3-1	2	7,507,474	71,197,424
+mTRP53-extended_(n=6)	3	2-2/3-1	3	6,270,685	61,929,404
+mTRP53-extended_(n=6)	3	3-2	1	2,238,850	40,003,596
+mTRP53-extended_(n=6)	3	3-2	2	1,868,744	36,313,404
+mTRP53-extended_(n=6)	3	3-2	3	1,641,661	26,337,938
Neuron survival					
+mGDF11_(n=6)	2	2-1	1	824,876	138,116
+mGDF11_(n=6)	2	2-1	2	1,060,304	167,050
+mGDF11_(n=6)	2	2-1	3	1,150,555	157,637
+mGDF11_(n=6)	2	2-2/3-1	1	2,617,892	233,651
+mGDF11_(n=6)	2	2-2/3-1	2	3,108,612	240,967
+mGDF11_(n=6)	2	2-2/3-1	3	2,572,953	198,165
+mGDNF+60bp (+SV40)_(n=18)	1	1-0	1	5,603,662	23,286,982
+mGDNF+60bp (+SV40)_(n=18)	1	1-0	2	4,184,035	17,924,162
+mGDNF+60bp (+SV40)_(n=18)	1	1-0	3	3,856,008	17,288,910
+mGDNF+60bp (+SV40)_(n=18)	1	1-1	1	4,440,121	15,561,043

+mGDNF+60bp (+SV40)_(n=18)	1	1-1	2	3,523,944	13,534,470
+mGDNF+60bp (+SV40)_(n=18)	1	1-1	3	3,400,972	13,790,621
+mGDNF+60bp (+SV40)_(n=18)	1	1-2	1	3,877,787	20,252,880
+mGDNF+60bp (+SV40)_(n=18)	1	1-2	2	2,673,036	10,192,005
+mGDNF+60bp (+SV40)_(n=18)	1	1-2	3	2,826,001	12,999,661
+mGDNF+60bp (+SV40)_(n=18)	1	1-3	1	853,883	7,146,307
+mGDNF+60bp (+SV40)_(n=18)	1	1-3	2	947,655	5,612,355
+mGDNF+60bp (+SV40)_(n=18)	1	1-3	3	1,188,121	9,335,436
+mGDNF+60bp (+SV40)_(n=18)	2	2-1	1	612,495	5,454,804
+mGDNF+60bp (+SV40)_(n=18)	2	2-1	2	757,718	5,783,183
+mGDNF+60bp (+SV40)_(n=18)	2	2-1	3	863,246	7,790,051
+mGDNF+60bp (+SV40)_(n=18)	2	2-2/3-1	1	5,028,986	31,286,586
+mGDNF+60bp (+SV40)_(n=18)	2	2-2/3-1	2	5,284,666	31,729,904
+mGDNF+60bp (+SV40)_(n=18)	2	2-2/3-1	3	5,067,256	29,722,294
+mGDNF+60bp (no_SV40)_(n=6)	2	2-1	1	647,219	4,508,758
+mGDNF+60bp (no_SV40)_(n=6)	2	2-1	2	1,046,806	5,580,139
+mGDNF+60bp (no_SV40)_(n=6)	2	2-1	3	1,084,279	5,640,239
+mGDNF+60bp (no_SV40)_(n=6)	2	2-2/3-1	1	7,411,704	30,738,986
+mGDNF+60bp (no_SV40)_(n=6)	2	2-2/3-1	2	6,393,391	25,575,520
+mGDNF+60bp (no_SV40)_(n=6)	2	2-2/3-1	3	6,137,552	25,982,916
+mPARK2-1_(n=9)	1	1-1	1	3,543,028	16,935,514
+mPARK2-1_(n=9)	1	1-1	2	3,333,414	17,952,554
+mPARK2-1_(n=9)	1	1-1	3	3,351,118	17,610,430
+mPARK2-1_(n=9)	1	1-2	1	3,136,001	18,941,298
+mPARK2-1_(n=9)	1	1-2	2	2,802,088	17,068,896
+mPARK2-1_(n=9)	1	1-2	3	3,059,956	20,676,398
+mPARK2-1_(n=9)	1	1-3	1	927,440	5,129,966
+mPARK2-1_(n=9)	1	1-3	2	902,328	4,799,656
+mPARK2-1_(n=9)	1	1-3	3	952,239	5,077,141

Co-transfection ratio testing					
(+SV40)-1:10_(n=3)	3	3-2	1	1,645,404	287,570,208
(+SV40)-1:10_(n=3)	3	3-2	2	1,930,140	371,473,824
(+SV40)-1:10_(n=3)	3	3-2	3	1,974,338	361,010,368
(+SV40)-1:25_(n=3)	3	3-2	1	530,215	207,639,472
(+SV40)-1:25_(n=3)	3	3-2	2	712,376	274,280,640
(+SV40)-1:25_(n=3)	3	3-2	3	870,615	346,753,056
+mAPP-1:10_(n=3)	3	3-2	1	2,273,780	407,255,712
+mAPP-1:10_(n=3)	3	3-2	2	2,684,856	488,776,960
+mAPP-1:10_(n=3)	3	3-2	3	2,703,615	496,153,856
+mAPP-1:25_(n=3)	3	3-2	1	805,977	404,170,272
+mAPP-1:25_(n=3)	3	3-2	2	886,807	532,754,624
+mAPP-1:25_(n=3)	3	3-2	3	1,225,314	627,739,200
(+SV40)-1:10_(n=3)	4	4-0	1	1,386,951	288,300,800
(+SV40)-1:10_(n=3)	4	4-0	2	1,540,991	309,665,472
(+SV40)-1:10_(n=3)	4	4-0	3	1,370,880	272,871,584
+mAPP-1:10_(n=3)	4	4-0	1	423,263	87,011,336
+mAPP-1:10_(n=3)	4	4-0	2	372,327	71,338,968
+mAPP-1:10_(n=3)	4	4-0	3	372,449	70,860,264
(+SV40)-1:20_(n=3)	4	4-0	1	778,535	302,028,096
(+SV40)-1:20_(n=3)	4	4-0	2	673,021	289,435,040
(+SV40)-1:20_(n=3)	4	4-0	3	792,652	322,403,232
+mAPP-1:20_(n=3)	4	4-0	1	254,617	101,555,672
+mAPP-1:20_(n=3)	4	4-0	2	247,934	110,987,800
+mAPP-1:20_(n=3)	4	4-0	3	182,580	70,870,440
(+SV40)-1:40_(n=3)	4	4-0	1	299,200	245,055,648
(+SV40)-1:40_(n=3)	4	4-0	2	292,438	226,919,584
(+SV40)-1:40_(n=3)	4	4-0	3	333,379	247,150,544
+mAPP-1:40_(n=3)	4	4-0	1	103,458	84,155,504
+mAPP-1:40_(n=3)	4	4-0	2	104,358	85,078,416
+mAPP-1:40_(n=3)	4	4-0	3	93,571	70,775,232
(+SV40)-1:80_(n=3)	4	4-0	1	176,014	294,641,728
(+SV40)-1:80_(n=3)	4	4-0	2	131,747	219,011,664
(+SV40)-1:80_(n=3)	4	4-0	3	168,178	263,804,832
+mAPP-1:80_(n=3)	4	4-0	1	58,271	94,549,552
+mAPP-1:80_(n=3)	4	4-0	2	52,746	96,008,240
+mAPP-1:80_(n=3)	4	4-0	3	60,100	103,506,440
(+SV40)-1:160_(n=3)	4	4-0	1	108,718	383,419,840
(+SV40)-1:160_(n=3)	4	4-0	2	91,695	322,608,352
(+SV40)-1:160_(n=3)	4	4-0	3	89,471	302,562,720
+mAPP-1:160_(n=3)	4	4-0	1	31,871	109,909,992
+mAPP-1:160_(n=3)	4	4-0	2	28,428	88,463,200
+mAPP-1:160_(n=3)	4	4-0	3	28,100	91,902,000

Table 7. (Renilla/Firefly) luminometer unadjusted and day-adjusted measurement ratios calculated from Table 6's raw measurement data, rounded to two decimal places.

3′UTR	Batch	Day	Measurement	(Renilla/ Firefly)	Log2(Renilla /Firefly)	Day-adjusted Log2(Renilla/ Firefly)
Viral sequences						
(+SV40)_(n=24)	1	1-0	1	17.96	4.17	-0.90
(+SV40)_(n=24)	1	1-0	2	18.06	4.17	-0.90
(+SV40)_(n=24)	1	1-0	3	18.15	4.18	-0.89
(+SV40)_(n=24)	1	1-1	1	19.09	4.26	-0.73
(+SV40)_(n=24)	1	1-1	2	25.69	4.68	-0.30
(+SV40)_(n=24)	1	1-1	3	26.14	4.71	-0.28
(+SV40)_(n=24)	1	1-2	1	29.93	4.90	-0.80
(+SV40)_(n=24)	1	1-2	2	38.18	5.25	-0.45
(+SV40)_(n=24)	1	1-2	3	35.12	5.13	-0.57
(+SV40)_(n=24)	1	1-3	2	55.97	5.81	0.07
(+SV40)_(n=24)	1	1-3	3	54.75	5.77	0.04
(+SV40)_(n=24)	2	2-1	1	99.48	6.64	0.14
(+SV40)_(n=24)	2	2-1	2	106.92	6.74	0.25
(+SV40)_(n=24)	2	2-1	3	104.93	6.71	0.22
(+SV40)_(n=24)	2	2-2/3-1	1	51.15	5.68	-0.07
(+SV40)_(n=24)	2	2-2/3-1	2	51.91	5.70	-0.05
(+SV40)_(n=24)	2	2-2/3-1	3	49.21	5.62	-0.12
(+SV40)_(n=24)	3	2-2/3-1	1	62.56	5.97	0.22
(+SV40)_(n=24)	3	2-2/3-1	2	58.49	5.87	0.13
(+SV40)_(n=24)	3	2-2/3-1	3	59.83	5.90	0.16
(+SV40)_(n=24)	3	3-2	1	139.20	7.12	0.36
(+SV40)_(n=24)	3	3-2	2	154.28	7.27	0.51
(+SV40)_(n=24)	3	3-2	3	128.51	7.01	0.25
+WPRE+hGH_(n=6)	3	2-2/3-1	1	262.50	8.04	2.29
+WPRE+hGH_(n=6)	3	2-2/3-1	2	258.86	8.02	2.27
+WPRE+hGH_(n=6)	3	2-2/3-1	3	244.83	7.94	2.19
+WPRE+hGH_(n=6)	3	3-2	1	404.45	8.66	1.90
+WPRE+hGH_(n=6)	3	3-2	2	462.57	8.85	2.09
+WPRE+hGH_(n=6)	3	3-2	3	429.08	8.75	1.98
Flex-cassette testing						
+0lox-F(+SV40)_(n=12)	2	2-1	1	37.91	5.24	-1.25
+0lox-F(+SV40)_(n=12)	2	2-1	2	36.99	5.21	-1.29
+0lox-F(+SV40)_(n=12)	2	2-1	3	40.00	5.32	-1.17
+0lox-F(+SV40)_(n=12)	2	2-2/3-1	1	24.47	4.61	-1.13
+0lox-F(+SV40)_(n=12)	2	2-2/3-1	2	24.27	4.60	-1.14
+0lox-F(+SV40)_(n=12)	2	2-2/3-1	3	23.11	4.53	-1.21
+0lox-F(+SV40)_(n=12)	3	2-2/3-1	1	21.26	4.41	-1.33
+0lox-F(+SV40)_(n=12)	3	2-2/3-1	2	20.57	4.36	-1.38
+0lox-F(+SV40)_(n=12)	3	3-2	1	48.60	5.60	-1.16
+0lox-F(+SV40)_(n=12)	3	3-2	2	47.38	5.57	-1.19
+0lox-F(+SV40)_(n=12)	3	3-2	3	44.99	5.49	-1.27
+1lox-F(+SV40)_(n=6)	2	2-1	1	29.48	4.88	-1.61

+1lox-F(+SV40)_(n=6)	2	2-1	2	25.83	4.69	-1.80
+1lox-F(+SV40)_(n=6)	2	2-1	3	26.16	4.71	-1.79
+1lox-F(+SV40)_(n=6)	2	2-2/3-1	1	17.19	4.10	-1.64
+1lox-F(+SV40)_(n=6)	2	2-2/3-1	2	17.07	4.09	-1.65
+1lox-F(+SV40)_(n=6)	2	2-2/3-1	3	17.49	4.13	-1.61
+flex-F(+SV40)_(n=18)	1	1-0	1	21.07	4.40	-0.67
+flex-F(+SV40)_(n=18)	1	1-0	2	24.18	4.60	-0.47
+flex-F(+SV40)_(n=18)	1	1-0	3	20.43	4.35	-0.72
+flex-F(+SV40)_(n=18)	1	1-1	1	4.33	2.11	-2.87
+flex-F(+SV40)_(n=18)	1	1-1	2	4.63	2.21	-2.77
+flex-F(+SV40)_(n=18)	1	1-1	3	4.24	2.09	-2.90
+flex-F(+SV40)_(n=18)	1	1-2	1	7.57	2.92	-2.79
+flex-F(+SV40)_(n=18)	1	1-2	2	8.15	3.03	-2.68
+flex-F(+SV40)_(n=18)	1	1-2	3	7.64	2.93	-2.77
+flex-F(+SV40)_(n=18)	1	1-3	1	16.85	4.07	-1.66
+flex-F(+SV40)_(n=18)	1	1-3	2	15.48	3.95	-1.78
+flex-F(+SV40)_(n=18)	1	1-3	3	16.35	4.03	-1.71
+flex-F(+SV40)_(n=18)	2	2-1	1	23.10	4.53	-1.96
+flex-F(+SV40)_(n=18)	2	2-1	2	22.33	4.48	-2.01
+flex-F(+SV40)_(n=18)	2	2-1	3	18.24	4.19	-2.31
+flex-F(+SV40)_(n=18)	2	2-2/3-1	1	13.26	3.73	-2.01
+flex-F(+SV40)_(n=18)	2	2-2/3-1	2	13.83	3.79	-1.95
+flex-F(+SV40)_(n=18)	2	2-2/3-1	3	14.32	3.84	-1.90
+flex-F (no_SV40)_(n=6)	2	2-1	1	21.00	4.39	-2.10
+flex-F (no_SV40)_(n=6)	2	2-1	2	19.88	4.31	-2.18
+flex-F (no_SV40)_(n=6)	2	2-1	3	20.34	4.35	-2.15
+flex-F (no_SV40)_(n=6)	2	2-2/3-1	1	12.09	3.60	-2.15
+flex-F (no_SV40)_(n=6)	2	2-2/3-1	2	11.11	3.47	-2.27
+flex-F (no_SV40)_(n=6)	2	2-2/3-1	3	11.36	3.51	-2.24
+0lox-B(+SV40)_(n=6)	2	2-1	1	86.56	6.44	-0.06
+0lox-B(+SV40)_(n=6)	2	2-1	2	79.10	6.31	-0.19
+0lox-B(+SV40)_(n=6)	2	2-1	3	85.29	6.41	-0.08
+0lox-B(+SV40)_(n=6)	2	2-2/3-1	1	52.45	5.71	-0.03
+0lox-B(+SV40)_(n=6)	2	2-2/3-1	2	54.16	5.76	0.02
+0lox-B(+SV40)_(n=6)	2	2-2/3-1	3	50.78	5.67	-0.08
+1lox-B(+SV40)_(n=6)	2	2-1	1	30.27	4.92	-1.57
+1lox-B(+SV40)_(n=6)	2	2-1	2	28.77	4.85	-1.65
+1lox-B(+SV40)_(n=6)	2	2-1	3	27.32	4.77	-1.72
+1lox-B(+SV40)_(n=6)	2	2-2/3-1	1	17.84	4.16	-1.59
+1lox-B(+SV40)_(n=6)	2	2-2/3-1	2	18.92	4.24	-1.50
+1lox-B(+SV40)_(n=6)	2	2-2/3-1	3	19.89	4.31	-1.43
+flex-B(+SV40)_(n=18)	1	1-0	1	2.14	1.10	-3.98
+flex-B(+SV40)_(n=18)	1	1-0	2	1.96	0.97	-4.10

+flex-B(+SV40)_(n=18)	1	1-0	3	2.14	1.09	-3.98
+flex-B(+SV40)_(n=18)	1	1-1	1	2.79	1.48	-3.50
+flex-B(+SV40)_(n=18)	1	1-1	2	2.65	1.40	-3.58
+flex-B(+SV40)_(n=18)	1	1-1	3	1.76	0.82	-4.17
+flex-B(+SV40)_(n=18)	1	1-2	1	4.84	2.28	-3.43
+flex-B(+SV40)_(n=18)	1	1-2	2	5.40	2.43	-3.27
+flex-B(+SV40)_(n=18)	1	1-2	3	5.45	2.45	-3.26
+flex-B(+SV40)_(n=18)	1	1-3	1	5.97	2.58	-3.16
+flex-B(+SV40)_(n=18)	1	1-3	2	6.00	2.59	-3.15
+flex-B(+SV40)_(n=18)	1	1-3	3	7.12	2.83	-2.90
+flex-B(+SV40)_(n=18)	2	2-1	1	12.30	3.62	-2.87
+flex-B(+SV40)_(n=18)	2	2-1	2	13.27	3.73	-2.76
+flex-B(+SV40)_(n=18)	2	2-1	3	12.93	3.69	-2.80
+flex-B(+SV40)_(n=18)	2	2-2/3-1	1	9.29	3.22	-2.53
+flex-B(+SV40)_(n=18)	2	2-2/3-1	2	9.25	3.21	-2.53
+flex-B(+SV40)_(n=18)	2	2-2/3-1	3	8.40	3.07	-2.67
+flex-B (no_SV40)_(n=6)	2	2-1	1	1.63	0.71	-5.79
+flex-B (no_SV40)_(n=6)	2	2-1	2	1.60	0.68	-5.81
+flex-B (no_SV40)_(n=6)	2	2-1	3	1.60	0.68	-5.81
+flex-B (no_SV40)_(n=6)	2	2-2/3-1	1	0.78	-0.36	-6.10
+flex-B (no_SV40)_(n=6)	2	2-2/3-1	2	0.71	-0.49	-6.23
+flex-B (no_SV40)_(n=6)	2	2-2/3-1	3	0.79	-0.33	-6.08
+flex-B+mDNAJB6-3_(n=9)	1	1-2	1	0.26	-1.93	-7.64
+flex-B+mDNAJB6-3_(n=9)	1	1-2	2	0.25	-2.03	-7.73
+flex-B+mDNAJB6-3_(n=9)	1	1-2	3	0.23	-2.10	-7.81
+flex-B+mDNAJB6-3_(n=9)	1	1-3	1	0.90	-0.15	-5.88
+flex-B+mDNAJB6-3_(n=9)	1	1-3	2	0.98	-0.03	-5.77
+flex-B+mDNAJB6-3_(n=9)	1	1-3	3	0.83	-0.28	-6.01
+flex-B+mDNAJB6-3_(n=9)	3	2-2/3-1	1	0.68	-0.55	-6.29
+flex-B+mDNAJB6-3_(n=9)	3	2-2/3-1	2	0.67	-0.59	-6.33
+flex-B+mDNAJB6-3_(n=9)	3	2-2/3-1	3	0.65	-0.62	-6.37
Anti-aggregation						
+mDNAJB6-3_(n=15)	1	1-1	1	10.97	3.46	-1.53
+mDNAJB6-3_(n=15)	1	1-1	2	11.59	3.53	-1.45

+mDNAJB6-3_(n=15)	1	1-1	3	11.84	3.57	-1.42
+mDNAJB6-3_(n=15)	1	1-2	1	10.95	3.45	-2.26
+mDNAJB6-3_(n=15)	1	1-2	2	11.88	3.57	-2.14
+mDNAJB6-3_(n=15)	1	1-2	3	12.61	3.66	-2.05
+mDNAJB6-3_(n=15)	1	1-3	1	7.56	2.92	-2.82
+mDNAJB6-3_(n=15)	1	1-3	2	7.55	2.92	-2.82
+mDNAJB6-3_(n=15)	1	1-3	3	7.57	2.92	-2.81
+mDNAJB6-3_(n=15)	2	2-1	1	12.21	3.61	-2.89
+mDNAJB6-3_(n=15)	2	2-1	2	13.56	3.76	-2.73
+mDNAJB6-3_(n=15)	2	2-1	3	13.51	3.76	-2.74
+mDNAJB6-3_(n=15)	2	2-2/3-1	1	7.29	2.87	-2.88
+mDNAJB6-3_(n=15)	2	2-2/3-1	2	7.18	2.84	-2.90
+mDNAJB6-3_(n=15)	2	2-2/3-1	3	7.53	2.91	-2.83
+mDNAJB6-2_(n=6)	2	2-1	1	149.65	7.23	0.73
+mDNAJB6-2_(n=6)	2	2-1	2	157.61	7.30	0.81
+mDNAJB6-2_(n=6)	2	2-1	3	164.98	7.37	0.87
+mDNAJB6-2_(n=6)	2	2-2/3-1	1	73.89	6.21	0.46
+mDNAJB6-2_(n=6)	2	2-2/3-1	2	77.95	6.28	0.54
+mDNAJB6-2_(n=6)	2	2-2/3-1	3	80.84	6.34	0.59
+mDNAJB6-1_(n=21)	1	1-1	1	0.97	-0.04	-5.02
+mDNAJB6-1_(n=21)	1	1-1	2	0.98	-0.03	-5.01
+mDNAJB6-1_(n=21)	1	1-1	3	1.04	0.05	-4.93
+mDNAJB6-1_(n=21)	1	1-2	1	1.05	0.08	-5.63
+mDNAJB6-1_(n=21)	1	1-2	2	1.00	0.00	-5.71
+mDNAJB6-1_(n=21)	1	1-2	3	1.14	0.19	-5.52
+mDNAJB6-1_(n=21)	1	1-3	1	2.82	1.50	-4.24
+mDNAJB6-1_(n=21)	1	1-3	2	2.41	1.27	-4.46
+mDNAJB6-1_(n=21)	1	1-3	3	2.46	1.30	-4.44
+mDNAJB6-1_(n=21)	2	2-1	1	3.31	1.73	-4.77
+mDNAJB6-1_(n=21)	2	2-1	2	3.36	1.75	-4.74
+mDNAJB6-1_(n=21)	2	2-1	3	3.46	1.79	-4.70
+mDNAJB6-1_(n=21)	2	2-2/3-1	1	2.56	1.35	-4.39
+mDNAJB6-1_(n=21)	2	2-2/3-1	2	2.46	1.30	-4.44
+mDNAJB6-1_(n=21)	2	2-2/3-1	3	2.36	1.24	-4.50
+mDNAJB6-1_(n=21)	3	2-2/3-1	1	2.62	1.39	-4.36
+mDNAJB6-1_(n=21)	3	2-2/3-1	2	2.59	1.37	-4.37
+mDNAJB6-1_(n=21)	3	2-2/3-1	3	2.59	1.37	-4.37
+mDNAJB6-1_(n=21)	3	3-2	1	7.56	2.92	-3.84
+mDNAJB6-1_(n=21)	3	3-2	2	6.75	2.75	-4.01
+mDNAJB6-1_(n=21)	3	3-2	3	6.55	2.71	-4.05
+mDNAJB6-1-extended_(n=6)	3	2-2/3-1	1	2.02	1.02	-4.73
+mDNAJB6-1-extended_(n=6)	3	2-2/3-1	2	1.88	0.91	-4.84
+mDNAJB6-1-extended_(n=6)	3	2-2/3-1	3	1.73	0.79	-4.95
+mDNAJB6-1-extended_(n=6)	3	3-2	1	4.98	2.31	-4.45

+mDNAJB6-1-extended_(n=6)	3	3-2	2	5.13	2.36	-4.40
+mDNAJB6-1-extended_(n=6)	3	3-2	3	4.66	2.22	-4.54
+mDNAJC6_(n=9)	1	1-1	1	82.22	6.36	1.38
+mDNAJC6_(n=9)	1	1-1	2	80.44	6.33	1.35
+mDNAJC6_(n=9)	1	1-1	3	75.85	6.24	1.26
+mDNAJC6_(n=9)	1	1-2	1	83.06	6.38	0.67
+mDNAJC6_(n=9)	1	1-2	2	94.42	6.56	0.85
+mDNAJC6_(n=9)	1	1-2	3	63.22	5.98	0.27
+mDNAJC6_(n=9)	1	1-3	1	93.33	6.54	0.81
+mDNAJC6_(n=9)	1	1-3	2	65.23	6.03	0.29
+mDNAJC6_(n=9)	1	1-3	3	86.62	6.44	0.70
+mHSPA1A_(n=12)	1	1-0	1	10.61	3.41	-1.66
+mHSPA1A_(n=12)	1	1-0	2	11.65	3.54	-1.53
+mHSPA1A_(n=12)	1	1-0	3	11.52	3.53	-1.54
+mHSPA1A_(n=12)	1	1-1	1	16.02	4.00	-0.98
+mHSPA1A_(n=12)	1	1-1	2	18.68	4.22	-0.76
+mHSPA1A_(n=12)	1	1-1	3	20.19	4.34	-0.65
+mHSPA1A_(n=12)	1	1-2	1	27.56	4.78	-0.92
+mHSPA1A_(n=12)	1	1-2	2	28.41	4.83	-0.88
+mHSPA1A_(n=12)	1	1-2	3	27.75	4.79	-0.91
+mHSPA1A_(n=12)	1	1-3	1	28.34	4.82	-0.91
+mHSPA1A_(n=12)	1	1-3	2	25.14	4.65	-1.08
+mHSPA1A_(n=12)	1	1-3	3	27.66	4.79	-0.95
+mMITF_(n=6)	2	2-1	1	33.87	5.08	-1.41
+mMITF_(n=6)	2	2-1	2	39.72	5.31	-1.18
+mMITF_(n=6)	2	2-1	3	41.72	5.38	-1.11
+mMITF_(n=6)	2	2-2/3-1	1	30.66	4.94	-0.81
+mMITF_(n=6)	2	2-2/3-1	2	32.34	5.02	-0.73
+mMITF_(n=6)	2	2-2/3-1	3	31.50	4.98	-0.77
+mMME_(n=9)	1	1-1	1	4.22	2.08	-2.91
+mMME_(n=9)	1	1-1	2	4.65	2.22	-2.77
+mMME_(n=9)	1	1-1	3	4.28	2.10	-2.88
+mMME_(n=9)	1	1-2	1	6.25	2.64	-3.06
+mMME_(n=9)	1	1-2	2	5.06	2.34	-3.37
+mMME_(n=9)	1	1-2	3	5.98	2.58	-3.13
+mMME_(n=9)	1	1-3	1	4.35	2.12	-3.61
+mMME_(n=9)	1	1-3	2	4.46	2.16	-3.58
+mMME_(n=9)	1	1-3	3	4.41	2.14	-3.60
+mPCSK1N_(n=9)	1	1-1	1	18.04	4.17	-0.81
+mPCSK1N_(n=9)	1	1-1	2	20.79	4.38	-0.61
+mPCSK1N_(n=9)	1	1-2	1	31.39	4.97	-0.74
+mPCSK1N_(n=9)	1	1-2	2	31.23	4.96	-0.74
+mPCSK1N_(n=9)	1	1-2	3	31.61	4.98	-0.73
+mPCSK1N_(n=9)	1	1-3	1	25.63	4.68	-1.06
+mPCSK1N_(n=9)	1	1-3	2	26.52	4.73	-1.01
+mPCSK1N_(n=9)	1	1-3	3	24.81	4.63	-1.10
+mSCG5_(n=9)	1	1-1	1	10.24	3.36	-1.63

+mSCG5_(n=9)	1	1-1	2	11.14	3.48	-1.51
+mSCG5_(n=9)	1	1-1	3	10.32	3.37	-1.62
+mSCG5_(n=9)	1	1-2	1	21.97	4.46	-1.25
+mSCG5_(n=9)	1	1-2	2	19.43	4.28	-1.43
+mSCG5_(n=9)	1	1-2	3	20.19	4.34	-1.37
+mSCG5_(n=9)	1	1-3	1	27.77	4.80	-0.94
+mSCG5_(n=9)	1	1-3	2	28.60	4.84	-0.90
+mSCG5_(n=9)	1	1-3	3	28.93	4.85	-0.88
Aggregation proteins						
+mAPP_(n=18)	1	1-0	1	51.42	5.68	0.61
+mAPP_(n=18)	1	1-0	2	54.68	5.77	0.70
+mAPP_(n=18)	1	1-0	3	50.80	5.67	0.60
+mAPP_(n=18)	1	1-1	1	62.46	5.96	0.98
+mAPP_(n=18)	1	1-1	2	69.75	6.12	1.14
+mAPP_(n=18)	1	1-1	3	65.41	6.03	1.05
+mAPP_(n=18)	1	1-2	1	107.39	6.75	1.04
+mAPP_(n=18)	1	1-2	2	115.85	6.86	1.15
+mAPP_(n=18)	1	1-2	3	120.02	6.91	1.20
+mAPP_(n=18)	1	1-3	1	88.47	6.47	0.73
+mAPP_(n=18)	1	1-3	2	90.57	6.50	0.76
+mAPP_(n=18)	1	1-3	3	87.31	6.45	0.71
+mAPP_(n=18)	3	2-2/3-1	1	80.89	6.34	0.59
+mAPP_(n=18)	3	2-2/3-1	2	80.99	6.34	0.60
+mAPP_(n=18)	3	3-2	1	166.94	7.38	0.62
+mAPP_(n=18)	3	3-2	2	193.87	7.60	0.84
+mAPP_(n=18)	3	3-2	3	154.93	7.28	0.52
+mAPP-extended_(n=6)	3	2-2/3-1	1	78.47	6.29	0.55
+mAPP-extended_(n=6)	3	2-2/3-1	2	78.40	6.29	0.55
+mAPP-extended_(n=6)	3	2-2/3-1	3	83.35	6.38	0.64
+mAPP-extended_(n=6)	3	3-2	1	170.85	7.42	0.66
+mAPP-extended_(n=6)	3	3-2	2	153.17	7.26	0.50
+mAPP-extended_(n=6)	3	3-2	3	155.34	7.28	0.52
+mITM2B_(n=12)	1	1-0	1	11.24	3.49	-1.58
+mITM2B_(n=12)	1	1-0	2	10.82	3.44	-1.63
+mITM2B_(n=12)	1	1-0	3	10.15	3.34	-1.73
+mITM2B_(n=12)	1	1-1	1	13.79	3.79	-1.20
+mITM2B_(n=12)	1	1-1	2	16.46	4.04	-0.94
+mITM2B_(n=12)	1	1-1	3	19.29	4.27	-0.71
+mITM2B_(n=12)	1	1-2	1	16.61	4.05	-1.65
+mITM2B_(n=12)	1	1-2	2	15.14	3.92	-1.79
+mITM2B_(n=12)	1	1-2	3	16.53	4.05	-1.66
+mITM2B_(n=12)	1	1-3	1	21.56	4.43	-1.31
+mITM2B_(n=12)	1	1-3	2	25.02	4.64	-1.09
+mITM2B_(n=12)	1	1-3	3	22.55	4.50	-1.24
+mSNCA_(n=15)	1	1-1	1	31.63	4.98	0.00
+mSNCA_(n=15)	1	1-1	2	41.20	5.36	0.38
+mSNCA_(n=15)	1	1-1	3	42.17	5.40	0.41
+mSNCA_(n=15)	1	1-2	1	55.45	5.79	0.09

+mSNCA_(n=15)	1	1-2	2	54.95	5.78	0.07
+mSNCA_(n=15)	1	1-2	3	54.48	5.77	0.06
+mSNCA_(n=15)	1	1-3	1	50.51	5.66	-0.08
+mSNCA_(n=15)	1	1-3	2	52.03	5.70	-0.03
+mSNCA_(n=15)	1	1-3	3	66.55	6.06	0.32
+mSNCA_(n=15)	3	2-2/3-1	1	52.61	5.72	-0.03
+mSNCA_(n=15)	3	2-2/3-1	2	54.15	5.76	0.02
+mSNCA_(n=15)	3	2-2/3-1	3	52.93	5.73	-0.02
+mSNCA_(n=15)	3	3-2	1	71.38	6.16	-0.60
+mSNCA_(n=15)	3	3-2	2	83.45	6.38	-0.38
+mSNCA_(n=15)	3	3-2	3	76.04	6.25	-0.51
+mSNCA-extended_(n=6)	3	2-2/3-1	1	36.93	5.21	-0.54
+mSNCA-extended_(n=6)	3	2-2/3-1	2	36.90	5.21	-0.54
+mSNCA-extended_(n=6)	3	2-2/3-1	3	36.75	5.20	-0.54
+mSNCA-extended_(n=6)	3	3-2	1	64.69	6.02	-0.74
+mSNCA-extended_(n=6)	3	3-2	2	71.72	6.16	-0.60
+mSNCA-extended_(n=6)	3	3-2	3	64.85	6.02	-0.74
+mIAPP-extended_(n=6)	3	2-2/3-1	1	28.40	4.83	-0.92
+mIAPP-extended_(n=6)	3	2-2/3-1	2	27.50	4.78	-0.96
+mIAPP-extended_(n=6)	3	2-2/3-1	3	29.64	4.89	-0.85
+mIAPP-extended_(n=6)	3	3-2	2	59.19	5.89	-0.87
+mIAPP-extended_(n=6)	3	3-2	3	54.45	5.77	-0.99
+mPRNP-extended_(n=6)	3	2-2/3-1	2	6.56	2.71	-3.03
+mPRNP-extended_(n=6)	3	2-2/3-1	3	6.63	2.73	-3.01
+mPRNP-extended_(n=6)	3	3-2	1	13.03	3.70	-3.06
+mPRNP-extended_(n=6)	3	3-2	2	12.83	3.68	-3.08
+mPRNP-extended_(n=6)	3	3-2	3	14.17	3.82	-2.94
+mACTB-extended_(n=6)	3	2-2/3-1	2	89.57	6.48	0.74
+mACTB-extended_(n=6)	3	2-2/3-1	3	95.35	6.58	0.83
+mACTB-extended_(n=6)	3	3-2	1	172.96	7.43	0.67

+mACTB-extended_(n=6)	3	3-2	2	169.62	7.41	0.65
+mACTB-extended_(n=6)	3	3-2	3	174.97	7.45	0.69
+mSOD1-extended_(n=6)	3	2-2/3-1	1	13.01	3.70	-2.04
+mSOD1-extended_(n=6)	3	2-2/3-1	2	13.62	3.77	-1.98
+mSOD1-extended_(n=6)	3	2-2/3-1	3	13.00	3.70	-2.04
+mSOD1-extended_(n=6)	3	3-2	1	29.97	4.91	-1.85
+mSOD1-extended_(n=6)	3	3-2	2	28.43	4.83	-1.93
+mSOD1-extended_(n=6)	3	3-2	3	29.12	4.86	-1.90
+mTRP53-extended_(n=6)	3	2-2/3-1	1	9.85	3.30	-2.44
+mTRP53-extended_(n=6)	3	2-2/3-1	2	9.48	3.25	-2.50
+mTRP53-extended_(n=6)	3	2-2/3-1	3	9.88	3.30	-2.44
+mTRP53-extended_(n=6)	3	3-2	1	17.87	4.16	-2.60
+mTRP53-extended_(n=6)	3	3-2	2	19.43	4.28	-2.48
+mTRP53-extended_(n=6)	3	3-2	3	16.04	4.00	-2.76
Neuron survival						
+mGDF11_(n=6)	2	2-1	1	0.17	-2.58	-9.07
+mGDF11_(n=6)	2	2-1	2	0.16	-2.67	-9.16
+mGDF11_(n=6)	2	2-1	3	0.14	-2.87	-9.36
+mGDF11_(n=6)	2	2-2/3-1	1	0.09	-3.49	-9.23
+mGDF11_(n=6)	2	2-2/3-1	2	0.08	-3.69	-9.43
+mGDF11_(n=6)	2	2-2/3-1	3	0.08	-3.70	-9.44
+mGDNF+60bp (+SV40)_(n=18)	1	1-0	1	4.16	2.06	-3.02
+mGDNF+60bp (+SV40)_(n=18)	1	1-0	2	4.28	2.10	-2.97
+mGDNF+60bp (+SV40)_(n=18)	1	1-0	3	4.48	2.16	-2.91
+mGDNF+60bp (+SV40)_(n=18)	1	1-1	1	3.50	1.81	-3.17
+mGDNF+60bp (+SV40)_(n=18)	1	1-1	2	3.84	1.94	-3.04
+mGDNF+60bp (+SV40)_(n=18)	1	1-1	3	4.05	2.02	-2.96
+mGDNF+60bp (+SV40)_(n=18)	1	1-2	1	5.22	2.38	-3.32

+mGDNF+60bp (+SV40)_(n=18)	1	1-2	2	3.81	1.93	-3.78
+mGDNF+60bp (+SV40)_(n=18)	1	1-2	3	4.60	2.20	-3.51
+mGDNF+60bp (+SV40)_(n=18)	1	1-3	1	8.37	3.07	-2.67
+mGDNF+60bp (+SV40)_(n=18)	1	1-3	2	5.92	2.57	-3.17
+mGDNF+60bp (+SV40)_(n=18)	1	1-3	3	7.86	2.97	-2.76
+mGDNF+60bp (+SV40)_(n=18)	2	2-1	1	8.91	3.15	-3.34
+mGDNF+60bp (+SV40)_(n=18)	2	2-1	2	7.63	2.93	-3.56
+mGDNF+60bp (+SV40)_(n=18)	2	2-1	3	9.02	3.17	-3.32
+mGDNF+60bp (+SV40)_(n=18)	2	2-2/3-1	1	6.22	2.64	-3.11
+mGDNF+60bp (+SV40)_(n=18)	2	2-2/3-1	2	6.00	2.59	-3.16
+mGDNF+60bp (+SV40)_(n=18)	2	2-2/3-1	3	5.87	2.55	-3.19
+mGDNF+60bp (no_SV40)_(n=6)	2	2-1	1	6.97	2.80	-3.69
+mGDNF+60bp (no_SV40)_(n=6)	2	2-1	2	5.33	2.41	-4.08
+mGDNF+60bp (no_SV40)_(n=6)	2	2-1	3	5.20	2.38	-4.12
+mGDNF+60bp (no_SV40)_(n=6)	2	2-2/3-1	1	4.15	2.05	-3.69
+mGDNF+60bp (no_SV40)_(n=6)	2	2-2/3-1	2	4.00	2.00	-3.74
+mGDNF+60bp (no_SV40)_(n=6)	2	2-2/3-1	3	4.23	2.08	-3.66
+mPARK2-1_(n=9)	1	1-1	1	4.78	2.26	-2.73
+mPARK2-1_(n=9)	1	1-1	2	5.39	2.43	-2.55
+mPARK2-1_(n=9)	1	1-1	3	5.26	2.39	-2.59
+mPARK2-1_(n=9)	1	1-2	1	6.04	2.59	-3.11
+mPARK2-1_(n=9)	1	1-2	2	6.09	2.61	-3.10
+mPARK2-1_(n=9)	1	1-2	3	6.76	2.76	-2.95
+mPARK2-1_(n=9)	1	1-3	1	5.53	2.47	-3.27
+mPARK2-1_(n=9)	1	1-3	2	5.32	2.41	-3.32
+mPARK2-1_(n=9)	1	1-3	3	5.33	2.41	-3.32
Co-transfection ratio testing						
(+SV40)-1:10_(n=3)	3	3-2	1	174.77	7.45	0.69
(+SV40)-1:10_(n=3)	3	3-2	2	192.46	7.59	0.83
(+SV40)-1:10_(n=3)	3	3-2	3	182.85	7.51	0.75
(+SV40)-1:25_(n=3)	3	3-2	1	391.61	8.61	1.85
(+SV40)-1:25_(n=3)	3	3-2	2	385.02	8.59	1.83

(+SV40)-1:25_(n=3)	3	3-2	3	398.29	8.64	1.88
+mAPP-1:10_(n=3)	3	3-2	1	179.11	7.48	0.72
+mAPP-1:10_(n=3)	3	3-2	2	182.05	7.51	0.75
+mAPP-1:10_(n=3)	3	3-2	3	183.51	7.52	0.76
+mAPP-1:25_(n=3)	3	3-2	1	501.47	8.97	2.21
+mAPP-1:25_(n=3)	3	3-2	2	600.76	9.23	2.47
+mAPP-1:25_(n=3)	3	3-2	3	512.31	9.00	2.24
(+SV40)-1:10_(n=3)	4	4-0	1	207.87	7.70	N/A
(+SV40)-1:10_(n=3)	4	4-0	2	200.95	7.65	N/A
(+SV40)-1:10_(n=3)	4	4-0	3	199.05	7.64	N/A
+mAPP-1:10_(n=3)	4	4-0	1	205.57	7.68	N/A
+mAPP-1:10_(n=3)	4	4-0	2	191.60	7.58	N/A
+mAPP-1:10_(n=3)	4	4-0	3	190.25	7.57	N/A
(+SV40)-1:20_(n=3)	4	4-0	1	387.94	8.60	N/A
(+SV40)-1:20_(n=3)	4	4-0	2	430.05	8.75	N/A
(+SV40)-1:20_(n=3)	4	4-0	3	406.74	8.67	N/A
+mAPP-1:20_(n=3)	4	4-0	1	398.86	8.64	N/A
+mAPP-1:20_(n=3)	4	4-0	2	447.65	8.81	N/A
+mAPP-1:20_(n=3)	4	4-0	3	388.16	8.60	N/A
(+SV40)-1:40_(n=3)	4	4-0	1	819.04	9.68	N/A
(+SV40)-1:40_(n=3)	4	4-0	2	775.96	9.60	N/A
(+SV40)-1:40_(n=3)	4	4-0	3	741.35	9.53	N/A
+mAPP-1:40_(n=3)	4	4-0	1	813.43	9.67	N/A
+mAPP-1:40_(n=3)	4	4-0	2	815.26	9.67	N/A
+mAPP-1:40_(n=3)	4	4-0	3	756.38	9.56	N/A
(+SV40)-1:80_(n=3)	4	4-0	1	1,673.97	10.71	N/A
(+SV40)-1:80_(n=3)	4	4-0	2	1,662.37	10.70	N/A
(+SV40)-1:80_(n=3)	4	4-0	3	1,568.60	10.62	N/A
+mAPP-1:80_(n=3)	4	4-0	1	1,622.58	10.66	N/A
+mAPP-1:80_(n=3)	4	4-0	2	1,820.20	10.83	N/A
+mAPP-1:80_(n=3)	4	4-0	3	1,722.24	10.75	N/A
(+SV40)-1:160_(n=3)	4	4-0	1	3,526.74	11.78	N/A
(+SV40)-1:160_(n=3)	4	4-0	2	3,518.28	11.78	N/A
(+SV40)-1:160_(n=3)	4	4-0	3	3,381.68	11.72	N/A
+mAPP-1:160_(n=3)	4	4-0	1	3,448.59	11.75	N/A
+mAPP-1:160_(n=3)	4	4-0	2	3,111.83	11.60	N/A
+mAPP-1:160_(n=3)	4	4-0	3	3,270.53	11.68	N/A